

Encyclopedia of  
Complexity and Systems Science Series  
*Editor-in-Chief: Robert A. Meyers*

SPRINGER NATURE  
Reference



Giuseppe Gaeta  
*Editor*

# Perturbation Theory

Mathematics, Methods and Applications

A Volume in the Encyclopedia of  
Complexity and Systems Science,  
Second Edition

---

# **Encyclopedia of Complexity and Systems Science Series**

**Editor-in-Chief**

Robert A. Meyers

The Encyclopedia of Complexity and Systems Science series of topical volumes provides an authoritative source for understanding and applying the concepts of complexity theory, together with the tools and measures for analyzing complex systems in all fields of science and engineering. Many phenomena at all scales in science and engineering have the characteristics of complex systems, and can be fully understood only through the transdisciplinary perspectives, theories, and tools of self-organization, synergetics, dynamical systems, turbulence, catastrophes, instabilities, nonlinearity, stochastic processes, chaos, neural networks, cellular automata, adaptive systems, genetic algorithms, and so on. Examples of near-term problems and major unknowns that can be approached through complexity and systems science include: The structure, history and future of the universe; the biological basis of consciousness; the integration of genomics, proteomics and bioinformatics as systems biology; human longevity limits; the limits of computing; sustainability of human societies and life on earth; predictability, dynamics and extent of earthquakes, hurricanes, tsunamis, and other natural disasters; the dynamics of turbulent flows; lasers or fluids in physics, microprocessor design; macromolecular assembly in chemistry and biophysics; brain functions in cognitive neuroscience; climate change; ecosystem management; traffic management; and business cycles. All these seemingly diverse kinds of phenomena and structure formation have a number of important features and underlying structures in common. These deep structural similarities can be exploited to transfer analytical methods and understanding from one field to another. This unique work will extend the influence of complexity and system science to a much wider audience than has been possible to date.

---

Giuseppe Gaeta  
Editor

# Perturbation Theory

Mathematics, Methods and  
Applications

A Volume in the Encyclopedia of Complexity  
and Systems Science, Second Edition

With 99 Figures and 4 Tables

 Springer



*Editor*

Giuseppe Gaeta  
Dipartimento di Matematica  
Università degli Studi di Milano  
Milano, Italy

ISSN 2629-2327                      ISSN 2629-2343 (electronic)  
Encyclopedia of Complexity and Systems Science Series  
ISBN 978-1-0716-2620-7              ISBN 978-1-0716-2621-4 (eBook)  
<https://doi.org/10.1007/978-1-0716-2621-4>

© Springer Science+Business Media, LLC, part of Springer Nature 2022

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

---

## Series Preface

The Encyclopedia of Complexity and System Science Series is a multivolume authoritative source for understanding and applying the basic tenets of complexity and systems theory as well as the tools and measures for analyzing complex systems in science, engineering, and many areas of social, financial, and business interactions. It is written for an audience of advanced university undergraduate and graduate students, professors, and professionals in a wide range of fields who must manage complexity on scales ranging from the atomic and molecular to the societal and global.

Complex systems are systems that comprise many interacting parts with the ability to generate a new quality of collective behavior through selforganization, e.g., the spontaneous formation of temporal, spatial, or functional structures. They are therefore adaptive as they evolve and may contain self-driving feedback loops. Thus, complex systems are much more than a sum of their parts. Complex systems are often characterized as having extreme sensitivity to initial conditions as well as emergent behavior that are not readily predictable or even completely deterministic. The conclusion is that a reductionist (bottom-up) approach is often an incomplete description of a phenomenon. This recognition that the collective behavior of the whole system cannot be simply inferred from the understanding of the behavior of the individual components has led to many new concepts and sophisticated mathematical and modeling tools for application to many scientific, engineering, and societal issues that can be adequately described only in terms of complexity and complex systems.

Examples of Grand Scientific Challenges which can be approached through complexity and systems science include: the structure, history, and future of the universe; the biological basis of consciousness; the true complexity of the genetic makeup and molecular functioning of humans (genetics and epigenetics) and other life forms; human longevity limits; unification of the laws of physics; the dynamics and extent of climate change and the effects of climate change; extending the boundaries of and understanding the theoretical limits of computing; sustainability of life on the earth; workings of the interior of the earth; predictability, dynamics, and extent of earthquakes, tsunamis, and other natural disasters; dynamics of turbulent flows and the motion of granular materials; the structure of atoms as expressed in the Standard Model and the formulation of the Standard Model and gravity into a Unified Theory; the

structure of water; control of global infectious diseases; and also evolution and quantification of (ultimately) human cooperative behavior in politics, economics, business systems, and social interactions. In fact, most of these issues have identified nonlinearities and are beginning to be addressed with nonlinear techniques, e.g., human longevity limits, the Standard Model, climate change, earthquake prediction, workings of the earth's interior, natural disaster prediction, etc.

The individual complex systems mathematical and modeling tools and scientific and engineering applications that comprised the *Encyclopedia of Complexity and Systems Science* are being completely updated and the majority will be published as individual books edited by experts in each field who are eminent university faculty members.

The topics are as follows:

- Agent Based Modeling and Simulation
- Applications of Physics and Mathematics to Social Science
- Cellular Automata, Mathematical Basis of
- Chaos and Complexity in Astrophysics
- Climate Modeling, Global Warming, and Weather Prediction
- Complex Networks and Graph Theory
- Complexity and Nonlinearity in Autonomous Robotics
- Complexity in Computational Chemistry
- Complexity in Earthquakes, Tsunamis, and Volcanoes, and Forecasting and Early Warning of Their Hazards
- Computational and Theoretical Nanoscience
- Control and Dynamical Systems
- Data Mining and Knowledge Discovery
- Ecological Complexity
- Ergodic Theory
- Finance and Econometrics
- Fractals and Multifractals
- Game Theory
- Granular Computing
- Intelligent Systems
- Nonlinear Ordinary Differential Equations and Dynamical Systems
- Nonlinear Partial Differential Equations
- Percolation
- Perturbation Theory
- Probability and Statistics in Complex Systems
- Quantum Information Science
- Social Network Analysis
- Soft Computing
- Solitons
- Statistical and Nonlinear Physics
- Synergetics
- System Dynamics
- Systems Biology

Each entry in each of the Series books was selected and peer reviews organized by one of our university-based book Editors with advice and consultation provided by our eminent Board Members and the Editor-in-Chief. This level of coordination assures that the reader can have a level of confidence in the relevance and accuracy of the information far exceeding than that generally found on the World Wide Web. Accessibility is also a priority and for this reason each entry includes a glossary of important terms and a concise definition of the subject. In addition, we are pleased that the mathematical portions of our Encyclopedia have been selected by Math Reviews for indexing in MathSciNet. Also, ACM, the world's largest educational and scientific computing society, recognized our *Computational Complexity: Theory, Techniques, and Applications* book, which contains content taken exclusively from the *Encyclopedia of Complexity and Systems Science*, with an award as one of the notable Computer Science publications. Clearly, we have achieved prominence at a level beyond our expectations, but consistent with the high quality of the content!

Palm Desert, CA, USA  
December 2022

Robert A. Meyers  
Editor-in-Chief

---

## Volume Preface

The idea behind *perturbation theory* is that when we are not able to determine exact solutions to a given problem, we might be able to determine approximate solutions to our problem starting from solutions to an approximate version of the problem, amenable to exact treatment. Thus, in a way, we use exact solutions to an approximate problem to get approximate solutions to an exact problem.

It goes without saying that many mathematical problems met in realistic situations, in particular as soon as we leave the linear framework, are not exactly solvable – either for an inherent impossibility or for our insufficient skills. Thus, perturbation theory is often the only way to approach realistic nonlinear systems.

It is implicit in the very nature of perturbation theory that it can only work once a problem which is both solvable – one also says “integrable” and in some sense “near” to the original problem – can be identified (it should be mentioned in this respect that the issue of “how near is near enough” is a delicate one). Quite often, the integrable problem to be used as a starting point is a linear one – may be obtained as the first-order expansion around a trivial or however known solution – and nonlinear corrections can be computed term by term via a recursive procedure based on expansion in a small parameter (usually denoted as  $\varepsilon$  by tradition); the point is that at each stage of this procedure, one should only solve linear equations, so that the procedure can, at least in principle, be carried over up to any desired order. In practice, one is limited by time, computational power, and the increasing dimension of the linear systems to be solved.

But limitations are not only due to the limits of the humans – or the computers – performing the actual computations: in fact, some delicate points arise when one considers the convergence of the  $\varepsilon$  series involved in the computations and in the expression of the solutions obtained by perturbation theory.

These points – that is, the power of perturbation theory, its basic features and tools, and its limitations in particular with regard to convergence issues – are discussed in the entry ► [“Perturbation Theory”](#) by G. Gallavotti. This entry also stresses the role which problems originating in physics had in the development of perturbation theory; and this is not only in historical terms (the computation of planetary orbits) but also in more recent times through the work of Poincaré first and then via quantum theory.

The modern setting of perturbation theory was laid down by Poincaré and goes through the use of what is today known as Poincaré normal forms; these are a cornerstone of the whole theory and hence, implicitly or explicitly, of all the entries presented in this section of the Encyclopedia. But they are also discussed in detail, together with their application, in the entry ► [“Normal Forms in Perturbation Theory”](#) by H. Broer.

The latter deals with the general problem, that is, with evolution differential equations (dynamical systems) with no special structure; or in applications originating from physics or engineering, one is often dealing with systems that (within a certain approximation) preserve energy and can be written in Hamiltonian form. In this case, as emphasized by Birkhoff, one can more efficiently consider perturbations of the Hamiltonian rather than of the equations of motion (the advantage originating in the fact that the Hamiltonian is a scalar function, while the equations of motion are a system of  $2n$  equations in  $2n$  dimensions). The normal form approach for Hamiltonian systems, and more generally Hamiltonian perturbation theory, is discussed in the entry ► [“Hamiltonian Perturbation Theory \(and Transition to Chaos\)”](#) by H. Broer and H. Hanßmann. This also discusses the problem of transition – as some control parameter, often the energy, is varied – from the regular behavior of the unperturbed system to the chaotic (“turbulent” if we deal with fluid motion) behavior displayed by many relevant Hamiltonian as well as non-Hamiltonian systems.

As mentioned above, in all the matters connected with perturbation theory and its applications, convergence issues play an extremely important role. They are discussed in the entry ► [“Convergence of Perturbative Expansions”](#) by S. Walcher, both in the general case and for Hamiltonian systems. The interplay between perturbations – and more generally changes in some relevant parameter characterizing the system within a more general family of system – and qualitative (not only quantitative) changes in its behavior is of course of general interest not only in the “extreme” case of transition from integrable to chaotic behavior but also when the qualitative change in the behavior of the system is somehow more moderate. Such a change is also known as a *bifurcation*. Albeit there is no entry specifically devoted to these, the reader will note that the concept of bifurcation appears in many, if not most, of the entries.

The behavior of a “generically perturbed” system depends on what is meant by “generically.” In particular if we deal with an unperturbed system which has some degree of symmetry, this may be an “accidental” feature – maybe due to the specially simple nature of integrable systems such as the one chosen as an unperturbed one – but might also correspond to a requirement by the very problem we are modeling; this is often the case when we deal with problems of physical or engineering origin, just because the fundamental equations of physics have some degree of symmetry. The presence of symmetry can be quite helpful – for example, in reducing the effective degrees of freedom of a given problem – and should be taken into account in the perturbative expansion. Moreover, the perturbative expansions can be made to have some degree of symmetry which can be used in the solution of the resulting equations. These matters are discussed at length in

the entry ► [“Symmetry and Perturbation Theory in Non-linear Dynamics”](#) by G. Gaeta.

A special – but widely applicable and very interesting – framework for the occurrence of bifurcation is provided by systems exhibiting parametric resonance. This is, for example, the case for an ample class of coupled oscillator systems, which would per se suffice to guarantee the phenomenon is of special interest in applications, besides its theoretical appeal. The analysis of parametric resonance from the point of view of perturbation theory is discussed in the entry ► [“Perturbation Analysis of Parametric Resonance”](#) by F. Verhulst.

As mentioned above, the transition from fully regular (integrable) to chaotic behavior is discussed in general terms in the entry ► [“Hamiltonian Perturbation Theory \(and Transition to Chaos\)”](#) by H. Broer and H. Hanßmann. However, quite remarkably, in some cases a perturbation will only moderately destroy the integrable behavior. This should be meant in the following sense: integrable behavior is characterized by the fact that whatever the initial conditions of the system, we are able to predict its behavior after an arbitrary long time.

It may happen that albeit this is not true, we are still able to predict either (a) the arbitrarily long-time behavior for a dense subset of all the possible initial conditions or (b) the exponentially long-time behavior for a subset of full measure of possible initial conditions (usually those “sufficiently near” to the exactly integrable case).

In the first case, the meaning of the statement is that any possible initial condition is “near” to an initial condition leading to an integrable-type behavior over all times (which does not imply its behavior will be near to integrable over arbitrary times, but only for sufficiently small – albeit this “small” could be extremely long on human scale – times). This kind of situation is investigated by the KAM theory (named after the initials of Kolmogorov, Arnold, and Moser), discussed in the entry ► [“Kolmogorov-Arnold-Moser \(KAM\) Theory for Finite and Infinite Dimensional Systems”](#) by L. Chierchia and M. Procesi; as the title indicates, this entry also explores the infinite-dimensional case, that is, KAM theory for Partial Differential Equations.

In the second case, the statements about stable behavior are valid only for a finite (albeit exponentially long, hence again often extremely long on human scale) time but apply to an open set of initial conditions. This approach was taken by Nekhoroshev and is presently known – together with the results obtained in this direction – as Nekhoroshev theory; this is the subject of the entry ► [“Nekhoroshev Theory”](#) by L. Niederman.

As mentioned above, the problem of planetary motions was historically at the origin of perturbation theory, since Ancient Greece; actually more recent results, including those due to the work of Poincaré – and those embodied in KAM and Nekhoroshev theories – also have roots in celestial mechanics (albeit then being used in completely different fields, for example, in the study of electron motion in a crystal). The application of perturbation theory in celestial mechanics is a very active field of research and the subject of the entry ► [“Perturbation Theory in Celestial Mechanics”](#) by A. Celletti.

In this context, one often considers reduced problems where not all the planets are taken into account; this is the origin of the “three-body problem,”

the three bodies being, for example, the Sun, Jupiter, and the Earth or the Earth, the Moon, and an artificial satellite. Much effort has been recently devoted to the study of special solutions for the  $N$ -body problem (i.e.,  $N$  bodies mutually attracting via potential forces) after the discovery of remarkable special solutions – termed “choreographies” – in which the bodies move along one or few common trajectories. This theory has not yet found applications in concrete physical systems or technology, but on the one hand these special solutions provide an organizing center for general nonlinear dynamics, and on the other the applicative potential of such collective motions (say, in micro-devices) is rather obvious. The  $N$ -body problem and these special solutions are discussed in the entry ► [“n-Body Problem and Choreographies”](#) by S. Terracini.

Needless to say, when one is faced with actual Celestial Mechanics problems, it is extremely difficult to work analytically, even at the perturbative level, and in most cases one has to resort to either computer algebra or numerical computations. Discussing these aspects is the aim of the entry ► [“Computational Methods in Perturbation Theory”](#) by A. Jorba.

In all these discussions, the theme is that of perturbations of an integrable system (usually, a linear one); in finite dimensions, this means a system having at least as many nontrivial constants of motion (first integrals, conserved quantities) as degrees of freedom. But, an integrable system could have more conserved quantities, as it happens, for example, for fully resonant harmonic oscillators. These are of course rather special cases, and correspondingly perturbations of such cases will display a special behavior. This matter is discussed in detail in the entry ► [“Perturbation of Superintegrable Hamiltonian Systems”](#) by F. Fassò.

As mentioned above, superintegrability may be related to resonance. Albeit resonance is commonly believed to make things more difficult, this is mainly true of *near-resonance* (see for example the entries ► [“Normal Forms in Perturbation Theory”](#) and ► [“Kolmogorov-Arnold-Moser \(KAM\) Theory for Finite and Infinite Dimensional Systems”](#) by Broer and by Chierchia and Procesi); in several circumstances it may be better to consider as unperturbed system a *fully resonant* one, considering the small differences between such a system and the one actually at hand in the perturbation. This method is also known as the “frequency deviation” method, or also the “detuning” one. This approach is indeed discussed in the entry ► [“Perturbation Theory and the Method of Detuning”](#) by G. Pucacco.

The reader has probably noted that all these problems correspond to smooth conservative systems, or at least to perturbation of such systems. When this is relaxed – which is not appropriate in studying the motion of planetary objects but may be appropriate in a number of contexts – the situation is both different and less well understood. The entry ► [“Perturbation of Systems with Nilpotent Real Part”](#) by the late T. Gramchev studies perturbation of linear systems with a nilpotent linear part, which is the case for dissipative unperturbed systems. The entry ► [“Perturbation Theory for Non-smooth Systems”](#) by M.A. Teixeira discusses the case of nonsmooth systems, which is in particular the case – quite relevant in real-world engineering applications – of systems with impacts.

A different kind of nonsmooth system is one we are all very familiar with: that is, legged locomotion – also called walking. A mathematical description



of this looks like a curiosity if we are thinking of humans (or animals) walk; but it is of course essential when one sets himself the task of building a walking robot, or some device to help impaired humans to walk. As it often happens when facing real-world (and thus not highly idealized) problems, this very concrete task calls for a rather high-note mathematics. This is illustrated in the entry ► [“Exact and Perturbation Methods in the Dynamics of Legged Locomotion”](#) by O. Makarenkov.

It was also mentioned that *quantum theory* was another major source of motivation for the development of perturbation theory, both historically and still in recent times. As for the latter, one should first of all remark how a standard tool in quantum perturbation theory, that is, the technique of Feynman diagrams – or, more generally, diagrammatic expansions – was incorporated into classical perturbation theory only in relatively recent times. The use of diagrammatic expansions in classical perturbation theory is discussed in the entry ► [“Diagrammatic Methods in Classical Perturbation Theory”](#) by G. Gentile.

Apart from this, the knowledge of the perturbation-theoretic techniques developed in the framework of quantum theory – both in general and for the study of atoms and molecules – is of general interest, both as a source of inspiration for tackling problems in different contexts and for the intrinsic interest of microscopic systems; while well known to physicists, this theory is maybe less known to mathematicians and engineers. The entries provided in this section of the encyclopedia can be an excellent entry point for those not familiar with this theory.

In the entry ► [“Perturbation Theory in Quantum Mechanics”](#) by L. Picasso, L. Bracci, and E. D’Emilio, the general setting and results are described, together with some selected special topics; the role of symmetries – and hence degeneracies – within quantum perturbation theory is paramount and also discussed here.

The entry ► [“Perturbation Theory and Molecular Dynamics”](#) by G. Panati focuses instead on the specific aspects of the perturbative approach to the quantum dynamics of molecules; this is a remarkable example of how taking into account the separation between slow and fast degrees of freedom allows one to deal with seemingly intractable problems.

A bridge between quantum and classical perturbation theory is provided by the semiclassical case, corresponding to taking into account the smallness of the energy scale set by Planck’s constant  $\hbar$  with respect to the energy scale involved in many (macroscopic or mesoscopic) problems. This is discussed in the entry ► [“Semiclassical Perturbation Theory”](#) by A. Sacchetti.

The quantum framework is also very interesting in connection with bifurcation theory; in this framework the “qualitative changes in the dynamics” which characterize bifurcations correspond to the qualitative changes in the spectrum. This in turn is related to monodromy on the mathematical side and to the problem of an atom in crossed magnetic and electric fields on the physical side. These matters, strongly related to several of those mentioned above, are discussed in the entry ► [“Quantum Bifurcations”](#) by B. Zhilinskii.

A key role in perturbation theory (quantum and classical) is played by the *adiabatic theorem*, and this both from the conceptual point of view and for

practical applications. Here the word “adiabatic” refers to the setting in which some of the control parameters of the system are changed in time; when this change is very slow, one says it is adiabatical (the terminology originates in classical Thermodynamics, but its Greek roots adapt well to what happens also in this context). In Quantum Mechanics, one is especially interested in determining the spectrum of the Hamiltonian, and the corresponding eigenfunctions. In this context, the adiabatic theorems state that – under suitable non-degeneracy conditions, which in the end are quite weak, and in particular provided no quantum bifurcation takes place – one can “follow” the deformation of the spectrum under a change in the control parameters. In particular, if one has a cyclic change in these, the resulting eigenfunctions will go back, after one cycle, to the original ones up to a change in phase. This kind of results is discussed in depth in the entry ► [“Quantum Adiabatic Theorem”](#) by S. Teufel.

Needless to say, the same problem of convergence of perturbative expansions met in classical dynamics and dynamical systems (► [“Convergence of Perturbative Expansions”](#)) is present in the quantum theory. This is tersely surveyed in the entry ► [“Convergent Perturbative Expansion in Condensed Matter and Quantum Field Theory”](#) by V. Mastropietro.

The group of quantum-related entries is completed by another entry, dealing with the situation where the unperturbed system consists of an ensemble of free particles, and the perturbation is made of the interactions among these. This approach is discussed in the entry ► [“Correlation Corrections as a Perturbation to the Quasi-free Approximation in Many-Body Quantum Systems”](#) by N. Benedikter and C. Boccato.

From the mathematical point of view, in the quantum case one deals with a partial differential equation – the Schrodinger equation – rather than with a system of ordinary differential equations (it should be noted that when dealing with the spectrum only, one is actually not requiring to study the full set of solutions to the concerned PDE).

Needless to say, this is not the only case where one has to deal with PDEs in the applications of continuum mechanics providing a classical framework where one is obliged to deal with PDEs. Rigorous results in perturbation theory for PDEs are not at the same level as for ODEs (and the insight provided by the quantum case is henceforth especially valuable).

The situation for KAM theory for PDEs is discussed in the entry by Chierchia and Procesi mentioned above (► [“Kolmogorov-Arnold-Moser \(KAM\) Theory for Finite and Infinite Dimensional Systems”](#)); despite its paramount relevance, KAM theory does not exhaust the scope of perturbation theory, and one would like to extend other results holding for ODEs to the setting of PDEs.

Research in this direction is very active and faces rather difficult problems despite the progresses obtained in recent years. Some of these results, together with an overview of the field, are described in the entry ► [“Perturbation Theory for PDEs”](#) by D. Bambusi. Quite appropriately, this entry ends up

stating that in several applied fields a sound understanding of PDEs' rigorous perturbation theory would be relevant for applications, mentioning in particular the water wave problem, quantum mechanics, electromagnetic theory and magnetohydrodynamics, and elastodynamics.

In particular, the problem of water waves is the subject of two further contributions collected here. The first of these is rather abstract and focuses on KAM theory for water waves; this is the entry on ► [“Perturbation Theory for Water Waves”](#) by R. Montalto. The other entry dealing with water waves focuses on a very relevant specific phenomenon, that is, that of *rogue waves* (also called anomalous waves). These waves arise from modulational instability and are waves of anomalously large amplitude with respect to the surrounding waves, arising apparently from nowhere and disappearing without leaving any trace. Deep sea water is the first environment where rogue waves have been studied, and the very term “rogue wave” was first coined by oceanographers (a rogue wave in relatively calm sea, with regular waves a few meters high, can exceed the height of 20 m and be very dangerous); but they are present also in other quite different contexts, such as nonlinear optics, Bose-Einstein condensates, acoustic turbulence in superfluid Helium, and in other areas of physics. Similarly, modulational instability is also a central concept in nonlinear optics, and so on. All these physical contexts have a common mathematical description, that is, the Nonlinear Schrodinger (NLS) equation, and correspondingly the theory developed in one context is readily applicable – and applied – in different physical contexts. The theory of rogue waves is discussed – reporting very recent results – in the entry ► [“Periodic Rogue Waves and Perturbation Theory”](#) by P. Grinevich, P. Santini, and F. Coppini.

This could also be a convenient way to conclude this Introduction, but in this case the reader would unavoidably remain with the impression that perturbation theory is mainly dealing with nonlinear problems originating in Physics or Engineering. While this is historically the origin of the most striking developments of the theory – and the realm in which it proved most successful – such characterization is by no means a built-in restriction.

Perturbation theory can also deal effectively with problems originating in different fields and having a rather different mathematical formulation, such as those arising in certain fields of Biology (beside fields where the mathematical formulation is anyway in terms of dynamical systems). This is shown concretely in the entry ► [“Perturbation of Equilibria in the Mathematical Theory of Evolution”](#) by A. Sanchez; in this case the problem is formulated in terms of evolutionary game theory. It should be noted that this is interesting not only for the intrinsic interest of the Darwin's theory of Evolution but also because game theory is increasingly used in rather diverse contexts.

The present volume represents the second edition of the section devoted to Perturbation Theory in the Encyclopedia of Complexity and Systems Science. It may be worth mentioning that in this second edition one-third of the entries are completely new; some of the entries contained in the first edition have also been updated and/or upgraded.

Finally, I would like to warmly thank all the authors for providing the remarkable entries making up this volume, as well as the referees who checked them anonymously and in several cases gave suggestions leading to improvements.

Milan, Italy  
December 2022

Giuseppe Gaeta  
Volume Editor

---

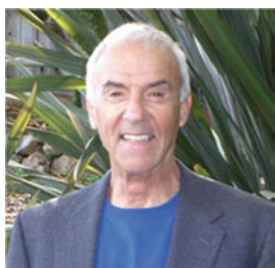
# Contents

|                                                                                                   |     |
|---------------------------------------------------------------------------------------------------|-----|
| <b>Perturbation Theory</b> .....                                                                  | 1   |
| Giovanni Gallavotti                                                                               |     |
| <b>Hamiltonian Perturbation Theory<br/>(and Transition to Chaos)</b> .....                        | 15  |
| Henk W. Broer and Heinz Hanßmann                                                                  |     |
| <b>Perturbation Theory in Quantum Mechanics</b> .....                                             | 47  |
| Luigi E. Picasso, Luciano Bracci, and Emilio d’Emilio                                             |     |
| <b>Normal Forms in Perturbation Theory</b> .....                                                  | 79  |
| Henk W. Broer                                                                                     |     |
| <b>Convergence of Perturbative Expansions</b> .....                                               | 105 |
| Sebastian Walcher                                                                                 |     |
| <b>Diagrammatic Methods in Classical Perturbation Theory</b> .....                                | 119 |
| Guido Gentile                                                                                     |     |
| <b>Perturbation Theory and the Method of Detuning</b> .....                                       | 141 |
| Giuseppe Pucacco                                                                                  |     |
| <b>Computational Methods in Perturbation Theory</b> .....                                         | 153 |
| Àngel Jorba                                                                                       |     |
| <b>Perturbation Analysis of Parametric Resonance</b> .....                                        | 167 |
| Ferdinand Verhulst                                                                                |     |
| <b>Symmetry and Perturbation Theory in Non-linear Dynamics</b> ....                               | 185 |
| Giuseppe Gaeta                                                                                    |     |
| <b>Perturbation of Systems with Nilpotent Real Part</b> .....                                     | 211 |
| Todor Gramchev                                                                                    |     |
| <b>Perturbation Theory for PDEs</b> .....                                                         | 229 |
| Dario Bambusi                                                                                     |     |
| <b>Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite<br/>Dimensional Systems</b> ..... | 247 |
| Luigi Chierchia and Michela Procesi                                                               |     |
| <b>Nekhoroshev Theory</b> .....                                                                   | 291 |
| Laurent Niederman                                                                                 |     |

|                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Perturbation of Superintegrable Hamiltonian Systems</b> .....                                                        | 307 |
| Francesco Fassò                                                                                                         |     |
| <b>Perturbation Theory in Celestial Mechanics</b> .....                                                                 | 339 |
| Alessandra Celletti                                                                                                     |     |
| <b>n-Body Problem and Choreographies</b> .....                                                                          | 357 |
| Susanna Terracini                                                                                                       |     |
| <b>Semiclassical Perturbation Theory</b> .....                                                                          | 391 |
| Andrea Sacchetti                                                                                                        |     |
| <b>Perturbation Theory and Molecular Dynamics</b> .....                                                                 | 409 |
| Gianluca Panati                                                                                                         |     |
| <b>Quantum Adiabatic Theorem</b> .....                                                                                  | 419 |
| Stefan Teufel                                                                                                           |     |
| <b>Quantum Bifurcations</b> .....                                                                                       | 433 |
| Boris Zhilinskiĭ                                                                                                        |     |
| <b>Convergent Perturbative Expansion in Condensed Matter and<br/>Quantum Field Theory</b> .....                         | 457 |
| Vieri Mastropietro                                                                                                      |     |
| <b>Correlation Corrections as a Perturbation to the Quasi-free<br/>Approximation in Many-Body Quantum Systems</b> ..... | 465 |
| Niels Benedikter and Chiara Boccato                                                                                     |     |
| <b>Perturbation of Equilibria in the Mathematical<br/>Theory of Evolution</b> .....                                     | 489 |
| Angel Sánchez                                                                                                           |     |
| <b>Perturbation Theory for Non-smooth Systems</b> .....                                                                 | 503 |
| Marco Antônio Teixeira                                                                                                  |     |
| <b>Exact and Perturbation Methods in the Dynamics of Legged<br/>Locomotion</b> .....                                    | 519 |
| Oleg Makarenkov                                                                                                         |     |
| <b>Perturbation Theory for Water Waves</b> .....                                                                        | 541 |
| Riccardo Montalto                                                                                                       |     |
| <b>Periodic Rogue Waves and Perturbation Theory</b> .....                                                               | 565 |
| F. Coppini, P. G. Grinevich, and P. M. Santini                                                                          |     |
| <b>Index</b> .....                                                                                                      | 585 |

---

## About the Editor-in-Chief



### **Robert A. Meyers**

President: RAMTECH Limited  
Manager, Chemical Process Technology,  
TRW Inc.  
Postdoctoral Fellow: California Institute of  
Technology  
Ph.D. Chemistry, University of California at  
Los Angeles  
B.A. Chemistry, California State University,  
San Diego

Dr. Meyers was manager of Energy and Environmental Projects at TRW (now Northrop Grumman) in Redondo Beach, CA, and is now president of RAMTECH Limited. He is coinventor of the Gravimelt process for desulfurization and demineralization of coal for air pollution and water pollution control and was manager of the Department of Energy project leading to the construction and successful operation of a first-of-a-kind Gravimelt Process Integrated Test Plant. Dr. Meyers is the inventor of and was project manager for the DOE-sponsored Magnetohydrodynamics Seed Regeneration Project which has resulted in the construction and successful operation of a pilot plant for production of potassium formate, a chemical utilized for plasma electricity generation and air pollution control. He also managed TRW efforts in magnetohydrodynamics electricity generating combustor and plasma channel development. Dr. Meyers managed the pilot-scale DoE project for determining the hydrodynamics of synthetic fuels. He is a coinventor of several thermooxidative stable polymers which have

achieved commercial success as the GE PEI, Upjohn Polyimides, and Rhone-Poulenc bismal-imide resins. He has also managed projects for photochemistry, chemical lasers, flue gas scrubbing, oil shale analysis and refining, petroleum analysis and refining, global change measurement from space satellites, analysis and mitigation (carbon dioxide and ozone), hydrometallurgical refining, soil and hazardous waste remediation, novel polymers synthesis, modeling of the economics of space transportation systems, space rigidizable structures, and chemiluminescence-based devices.

He is a senior member of the American Institute of Chemical Engineers, member of the American Physical Society, and member of the American Chemical Society and has served on the UCLA Chemistry Department Advisory Board. He was a member of the joint USA-Russia working group on air pollution control and the EPA-sponsored Waste Reduction Institute for Scientists and Engineers.

Dr. Meyers has more than 20 patents and 50 technical papers in the fields of photochemistry, pollution control, inorganic reactions, organic reactions, luminescence phenomena, and polymers. He has published in primary literature journals including *Science* and the *Journal of the American Chemical Society*, and is listed in *Who's Who in America* and *Who's Who in the World*. Dr. Meyers' scientific achievements have been reviewed in feature articles in the popular press in publications such as *The New York Times Science Supplement* and *The Wall Street Journal* as well as more specialized publications such as *Chemical Engineering and Coal Age*. A public service film was produced by the Environmental Protection Agency on Dr. Meyers' chemical desulfurization invention for air pollution control.

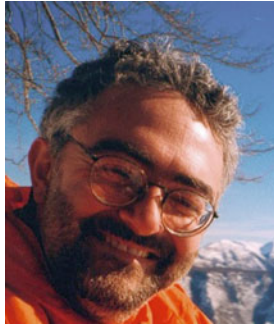
Dr. Meyers is the author or editor-in-chief of a wide range of technical books including the *Handbook of Chemical Production Processes*; the *Handbook of Synfuels Technology*; the *Handbook of Petroleum Refining Processes*, now in fourth edition; the *Handbook of Petrochemical Production Processes* (McGraw-Hill), now in a second edition; the *Handbook of Energy Technology and Economics*, published by John Wiley & Sons; *Coal Structure*, published by Academic Press; and *Coal Desulfurization* as well as the *Coal Handbook*



published by Marcel Dekker. He served as chairman of the advisory board for *A Guide to Nuclear Power Technology*, published by John Wiley & Sons, which won the Association of American Publishers Award as the best book in technology and engineering. He also served as editor-in-chief of three editions of the *Elsevier Encyclopedia of Physical Science and Technology*. Most recently, Dr. Meyers serves as editor-in-chief of the *Encyclopedia of Analytical Chemistry* as well as *Reviews in Cell Biology and Molecular Medicine* and a book series of the same name both published by John Wiley & Sons. In addition, Dr. Meyers currently serves as editor-in-chief of two Springer Nature book series, *Encyclopedia of Complexity and Systems Science* and *Encyclopedia of Sustainability Science and Technology*.

---

## About the Volume Editor



**G. Gaeta** studied Theoretical Physics in Roma and New York; he had postdoctoral positions in Montreal, Paris, Madrid, and Utrecht and was Reader in Nonlinear Systems in Loughborough, UK, before landing in Milano, Italy, where he is now Professor of Mathematical Physics. His main fields of research are Mathematical Physics – in particular Nonlinear Systems, Perturbation Theory, and Symmetry – and Mathematical Biology. He is the author of over 150 research papers and 3 research monographs, together with a textbook in introductory Mathematical Biology. He is also the main organizer of the SPT (Symmetry and Perturbation Theory) series of conferences, which started in 1996, and the Director of the virtual Research Institute SMRI.

---

## Contributors

**Dario Bambusi** Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

**Niels Benedikter** Dipartimento di Matematica ‘Federigo Enriques’, Università degli Studi di Milano, Milan, Italy

**Chiara Boccato** Dipartimento di Matematica ‘Federigo Enriques’, Università degli Studi di Milano, Milan, Italy

**Luciano Bracci** Dipartimento di Fisica, Università di Pisa, Pisa, Italy

**Henk W. Broer** Instituut voor Wiskunde en Informatica, Rijksuniversiteit Groningen, Groningen, The Netherlands

**Alessandra Celletti** Dipartimento di Matematica, Università di Roma Tor Vergata, Rome, Italy

**Luigi Chierchia** Dipartimento di Matematica e Fisica, Università “Roma Tre”, Rome, Italy

**F. Coppini** Dipartimento di Fisica, Università di Roma, “La Sapienza”, Rome, Italy

Department of Mathematics, Physics and Electrical Engineering, Northumbria University Newcastle, Newcastle upon Tyne, UK

**Emilio d’Emilio** Dipartimento di Fisica, Università di Pisa, Pisa, Italy

**Francesco Fassò** Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, Padova, Italy

**Giuseppe Gaeta** Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

**Giovanni Gallavotti** Dipartimento di Fisica and I.N.F.N., sezione Roma-I, Università di Roma I “La Sapienza”, Rome, Italy

**Guido Gentile** Dipartimento di Matematica, Università di Roma Tre, Rome, Italy

**Todor Gramchev** Dipartimento di Matematica e Informatica, Università di Cagliari, Cagliari, Italy

---

Todor Gramchev: deceased.

**P. G. Grinevich** Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia

L.D. Landau Institute for Theoretical Physics, Chernogolovka, Moscow region, Russia

Department of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia

**Heinz Hanßmann** Mathematisch Instituut, Universiteit Utrecht, Utrecht, The Netherlands

**Àngel Jorba** Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain

**Oleg Makarenkov** Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA

**Vieri Mastropietro** Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

**Riccardo Montalto** Dipartimento di Matematica “Federigo Enriques”, Università degli Studi di Milano, Milan, Italy

**Laurent Niederman** Topologie et Dynamique – UMR 8628 du CNRS, Université Paris, Paris, France

Astronomie et Systèmes Dynamiques – UMR 8028 du CNRS, IMCCE, Paris, France

**Gianluca Panati** Dipartimento di Matematica, Università di Roma “La Sapienza”, Rome, Italy

**Luigi E. Picasso** Dipartimento di Fisica, Università di Pisa, Pisa, Italy

**Michela Procesi** Dipartimento di Matematica e Fisica, Università “Roma Tre”, Rome, Italy

**Giuseppe Pucacco** Department of Physics, University of Rome Tor Vergata, Rome, Italy

**Andrea Sacchetti** Dipartimento di Matematica Pura ed Applicata, Università di Modena e Reggio Emilia, Modena, Italy

**Angel Sánchez** Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Madrid, Spain

Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, Spain

**P. M. Santini** Dipartimento di Fisica, Università di Roma, “La Sapienza”, Rome, Italy

**Marco Antônio Teixeira** Department of Mathematics, Universidade Estadual de Campinas, Campinas, Brazil

**Susanna Terracini** Dipartimento di Matematica “Giuseppe Peano”,  
Universita di Torino, Turin, Italy

**Stefan Teufel** Fachbereich Mathematik, Universität Tübingen, Tübingen,  
Germany

**Ferdinand Verhulst** Mathematisch Instituut, University of Utrecht, Utrecht,  
The Netherlands

**Sebastian Walcher** Lehrstuhl A für Mathematik, RWTH Aachen, Aachen,  
Germany

**Boris Zhilinskiĭ** Université du Littoral, Dunkerque, France

## Perturbation Theory

Giovanni Gallavotti

Dipartimento di Fisica and I.N.F.N., sezione  
Roma-I, Università di Roma I “La Sapienza”,  
Rome, Italy

### Article Outline

Glossary

Definition of the Subject

Introduction

Poincaré’s Theorem and Quanta

Mathematics and Physics. Renormalization

Need of Convergence Proofs

Multiscale Analysis

A Paradigmatic Example of PT Problem

Lindstedt Series

Convergence. Scales. Multiscale Analysis

Non Convergent Cases

Conclusion and Outlook

Future Directions

Bibliography

### Glossary

**Formal power series** A power series, giving the value of a function  $f(\varepsilon)$  of a parameter  $\varepsilon$ , that is derived assuming that  $f$  is analytic in  $\varepsilon$ .

**Renormalization group** Method for multiscale analysis and resummation of formal power series. Usually applied to define a systematic collection of terms to organize a formal power series into a convergent one.

**Lindstedt series** An algorithm to develop formal power series for computing the parametric equations of invariant tori in systems close to integrable.

**Multiscale problem** Any problem in which an infinite number of scales play a role.

## Definition of the Subject

**Perturbation Theory:** Computation of a quantity depending on a parameter  $\varepsilon$  starting from the knowledge of its value for  $\varepsilon = 0$  by deriving a power series expansion in  $\varepsilon$ , under the assumption of its existence, and if possible discussing the interpretation of the series. Perturbation theory is very often the only way to get a glimpse of the properties of systems whose equations cannot be “explicitly solved” in computable form.

The importance of Perturbation Theory is witnessed by its applications in Astronomy, where it led not only to the discovery of new planets (Neptune) but also to the discovery of Chaotic motions, with the completion of the Copernican revolution and the full understanding of the role of Aristotelian Physics formalized into uniform rotations of deferents and epicycles (today Fourier representation of quasi periodic motions). It also played an essential role in the development of Quantum Mechanics and the understanding of the periodic table. The successes of Quantum Field Theory in Electrodynamics first, then in Strong interactions and finally in the unification of the elementary forces (strong, electromagnetic, and weak) are also due to perturbation theory, which has also been essential in the theoretical understanding of the critical point universality. The latter two themes concern the new methods that have been developed in the last 50 years, marking a kind of new era for perturbation theory; namely dealing with singular problems, via the techniques called, in Physics, “Renormalization Group” and, in Mathematics, “Multiscale Analysis”.

### Introduction

Perturbation theory, henceforth PT, arises when the value of a function of interest is associated with a problem depending on a parameter, here

called  $\varepsilon$ . The value has to be a simple, or at least explicit and rigorous, computation for  $\varepsilon = 0$  while its computation for  $\varepsilon \neq 0$ , small, is attempted by expressing it as the sum of a power series in  $\varepsilon$  which will be called here the “solution”.

It is important to say since the beginning that a real PT solution of a problem involves two distinct steps: the first is to show that assuming that there is a convergent power series solving the problem then the *coefficients* of the  $n$ th power of  $\varepsilon$  exist and can be computed via finite computation. The resulting series will be called *formal solution* or formal series for the problem. The second step, that will be called *convergence theory*, is to prove that the formal series converges for  $\varepsilon$  small enough, or at least find a “summation rule” that gives a meaning to the formal series thus providing a real solution to the problem. None of the two problems is trivial, in the interesting cases, although the second is certainly the key and a difficult one.

Once Newton’s law of universal gravitation was established it became necessary to develop methods to find its implications. Laplace’s “*Mécanique Céleste*” (Laplace 1799), provided a detailed and meticulous exposition of a general method that has become a classic, if not the first, example of perturbation theory, quite different from the parallel analysis of Gauss which can be more appropriately considered a “non perturbative” development.

Since Laplace one can say that many applications along his lines followed. In the XIX century wide attention was dedicated to extend Laplace’s work to cover various astronomical problems: tables of the coefficients were dressed and published, and algorithms for their construction were devised, and planets were discovered (Neptune, 1846). Well known is the “Lindstedt algorithm” for the computation of the  $n$ th order coefficients of the PT series for the non resonant quasi periodic motions. The algorithm provides a power series representation for the quasi periodic motions with non resonant frequencies which is extremely simple: however it represents the  $n$ th coefficient as a sum of many terms, some of size of the order of a power on  $n!$ . Which of course is a serious problem for the convergence.

It became a central issue, known as the “small denominators problem” after Poincaré’s deep

critique of the PT method, generated by his analysis of the three-body problem. It led to his “non-integrability theorem” about the generic nonexistence of convergent power series in the perturbation parameter  $\varepsilon$  whose sum would be a constant of motion for a Hamiltonian  $H_\varepsilon$ , member of a family of Hamiltonians parametrized by  $\varepsilon$  and reducing to an integrable system for  $\varepsilon = 0$ . The theorem suggested (to some) that even the PT series of Lindstedt (to which Poincaré’s theorem does not apply) could be meaningless even though formally well defined (Poincaré 1892).

A posteriori, it should be recognized that PT was involved also in the early developments of Statistical Mechanics in the XIX century: the virial theorem application to obtain the Van der Waals equation of state can be considered a first-order calculation in PT (although this became clear only a century later with the identification of  $\varepsilon$  as the inverse of the space dimension).

## Poincaré’s Theorem and Quanta

With Poincaré begins a new phase: the question of convergence of series in  $\varepsilon$  becomes a central one in the Mathematics literature. Much less, however, in the Physics literature where the new discoveries in the atomic phenomena attracted the attention. It seems that in the Physics research it was taken for granted that convergence was not an issue: atomic spectra were studied via PT and early authoritative warnings were simply disregarded (for instance, explicit by Einstein, in (Einstein 1917), and clear, in (Fermi 1923), but “timid” being too far against the mainstream, for his young age). In this way quantum theory could grow from the original formulations of Bohr, Sommerfeld, Ehrenfest relying on PT to the final formulations of Heisenberg and Schrödinger quite far from it. Nevertheless, the triumph of quantum theory was quite substantially based on the technical development and refinement of the methods of formal PT: the calculation of the Compton scattering, the Lamb shift, Fermi’s weak interactions model and other spectacular successes came in spite of the parallel recognition that some of the series that

were being laboriously computed not only could not possibly be convergent but their very existence, to all orders  $n$ , was in doubt.

The later Feynman graphs representation of PT was a great new tool which superseded and improved earlier graphical representations of the calculations. Its simplicity allowed a careful analysis and understanding of cases in which even *formal* PT seemed puzzlingly failing.

Renormalization theory was developed to show that the convergence problems that seemed to plague even the computation of the individual coefficients of the series, hence the formal PT series at fixed order, were, in reality, often absent, in great generality, as suspected by the earlier treatments of special (important) cases, like the higher-order evaluations of the Compton scattering, and other quantum electrodynamics cross sections or anomalous characteristic constants (e.g., the magnetic moment of the muon).

## Mathematics and Physics. Renormalization

In 1943 the first important result on the convergence of the series of the Lindstedt kind was obtained by Siegel (Siegel 1943): a formal PT series, of interest in the theory of complex maps, was shown to be convergent. Siegel's work was certainly a stimulus for the later work of Kolmogorov who solved (Kolmogorov 1954), a problem that had been considered not soluble by many: to find the convergence conditions and the convergence proof of the Lindstedt series for the quasi periodic motions of a generic analytic Hamiltonian system, in spite of Poincaré's theorem and actually avoiding contradiction with it. Thus, showing the soundness of the comments about the unsatisfactory aspects of Poincaré's analysis that had been raised almost immediately by Weierstrass, Hadamard and others.

In 1956 not only Kolmogorov theorem appeared but also convergence of another well known and widely used formal series, the virial series, was achieved in an unnoticed work by Morrey (Morrey 1955), and independently rediscovered in the early 1960's.

At this time it seems that all series with well-defined terms were thought to be either convergent or at least asymptotic: for most Physicists convergence or asymptoticity were considered of little interest and matters to be left to Mathematicians.

However, with the understanding of the formal aspects of renormalization theory the interest in the convergence properties of the formal PT series once again became the center of attention.

On the one hand mathematical proofs of the existence of the PT series, for interesting quantum fields models, to all orders were investigated settling the question once and for all (Hepp's theorem (Hepp 1966)); on the other hand it was obvious that even if convergent (like in the virial or Meyer expansions, or in the Kolmogorov theory) it was well understood that the radius of convergence would not be large enough to cover all the physically interesting cases. The sum of the series would in general become singular in the sense of analytic functions and, even if admitting analytic continuation beyond the radius of convergence, a singularity in  $\varepsilon$  would be eventually hit. The singularity was supposed to correspond to very important phenomena, like the critical point in statistical mechanics or the onset of chaotic motions (already foreseen by Poincaré in connection with his non convergence theorem). Thus, research developed in two directions.

The first aimed at understanding the nature of the singularities from the formal series coefficients: in the 1960s many works achieved the understanding of the scaling laws (i.e., some properties of the divergences appearing at the singularities of the PT series or of its analytic continuation, for instance in the work of M. Fisher, Kadanoff, Widom and many others). This led to trying to find resummations, i.e., to collect terms of the formal series to transform them into *convergent series* in terms of *new parameters*, the running couplings.

The latter would be singular functions of the original  $\varepsilon$  thus possibly reducing the study of the singularity to the singularities of the running couplings. The latter could be studied by independent methods, typically by studying the iterations of an auxiliary dynamical system (called the *beta*



*function flow*). This was the *approach* or *renormalization group method* of Wilson (Gallavotti 2001; Wilson and Kogut 1973).

The second direction was dedicated to finding out the real meaning of the PT series in the cases in which convergence was doubtful or a priori excluded: in fact already Landau had advanced the idea that the series could be just illusions in important problems like the paradigmatic quantum field theory of a scalar field or the fundamental quantum electrodynamics (Frölich 1982; Gross 1999).

In a rigorous treatment the function that the series were supposed to represent would be in fact a trivial function with a dependence on  $\epsilon$  unrelated to the coefficients of the well defined and non trivial but formal series. It was therefore important to show that there were at least cases in which the perturbation series of a nontrivial problem had a meaning determined by its coefficients. This was studied in the scalar model of quantum field theory and a proof of “non triviality” was achieved after the ground-breaking work of Nelson on two-dimensional models (Nelson 1966; Simon 1974): soon followed by similar results in two dimensions and the difficult extension to three-dimensional models by Glimm and Jaffe (Glimm and Jaffe 1981), and generating many works and results on the subject which took the name of “constructive field theory” (Gallavotti 1985b).

But Landau’s triviality conjecture was actually dealing with the “real problem”, i.e., the 4-dimensional quantum fields. The conjecture remains such at the moment, in spite of very intensive work and attempts at its proof. The problem had relevance because it could have meant that not only the simple scalar models of constructive field theory were trivial but also the QED series which had received strong experimental support with the correct prediction of fine structure phenomena could be illusions, in spite of their well-defined PT series: which would remain as mirages of a non existing reality.

The work of Wilson made clear that the “triviality conjecture” of Landau could be applied only to theories which, after the mentioned resummations, would be controlled by a beta function flow that could not be studied perturbatively, and introduced

the new notion of asymptotic freedom. This is a property of the beta function flow, implying that the running couplings are bounded and small so that the resummed series are more likely to have a meaning (Wilson and Kogut 1973).

This work revived the interest in PT for quantum fields with attention devoted to new models that had been believed to be non renormalizable. Once more the apparently preliminary problem of developing a formal PT series played a key role: and it was discovered that many Yang–Mills quantum field theories were in fact renormalizable in the ultraviolet region (‘t Hooft 1999; ‘t Hooft and Veltman 1972), and an exciting period followed with attempts at using Wilson’s methods to give a meaning to the Yang–Mills theory with the hope of building a theory of the strong interactions. Thus, it was discovered that several Yang–Mills theories were asymptotically free as a consequence of the high symmetry of the model, proving that what seemed to be strong evidence that no renormalizable model would have asymptotic freedom was an ill-founded belief (that in a sense slowed down the process of understanding, and not only of the strong interactions).

Suddenly understanding the strong interactions, until then considered an impossible problem became possible (Gross 1999), as solutions could be written and *effectively computed* in terms of PT which, although not proved to be convergent or asymptotic (still an open problem in dimension  $d = 4$ ) were immune to the argument of Landau. The impact of the new developments led a little later to the unification of all interactions into the *standard model* for the theory of elementary particles (including the electromagnetic and weak interactions). The standard model was shown to be asymptotically free *even in the presence of symmetry breaking*, at least if a few other interactions in the model (for instance the Higgs particle self interaction) were treated heuristically while waiting for the discovery of the “Higgs particle” and for a better understanding of the structure of the elementary particles at length scales intermediate between the Fermi scale ( $\sim 10^{-15}$  cm (the weak interactions scale)) and the Planck scale (the gravitational interaction scale, 15 orders of magnitude below).

Given that the very discovery of renormalizability of Yang–Mills fields and the birth of a strong interactions theory had been firmly-grounded on experimental results (Gross 1999), the latter “missing step” was, and still is, considered an acceptable gap.

## Need of Convergence Proofs

The story of the standard model is paradigmatic of the power of PT: it should convince anyone that the analysis of formal series, including their representation by diagrams, which plays an essential part, is to be taken seriously. PT is certainly responsible for the revival and solution of problems considered by many as hopeless.

In a sense PT in the elementary particles domain can only, so far, partially be considered a success. Different is the situation in the developments that followed the works of Siegel and Kolmogorov. Their relevance for Celestial Mechanics and for several problems in applied physics (particle accelerator design, nuclear fusion machines for instance) and for statistical mechanics made them too the object of a large amount of research work.

The problems are simpler to formulate and often very well posed but the possibility of existence of chaotic motions, always looming, made it imperative not to be content with heuristic analysis and imposed the quest of mathematically complete studies. The lead were the works of Siegel and Kolmogorov. They had established convergence of certain PT series, but there were other series which would certainly be not convergent even though formally well defined and the question was, therefore, which would be their meaning.

More precisely it was clear that the series could be used to find approximate solutions to the equations, representing the motion for very long times under the assumption of “small enough”  $\varepsilon$ . But this could hardly be considered an understanding of the PT series in Mechanics: the estimated values of  $\varepsilon$  would have to be too small to be of interest, with the exception of a few special cases. The real question was what could be done to give the PT series the status of exact solution.

As we shall see the problem is deeply connected with the above-mentioned asymptotic freedom: this is perhaps not surprising because the link between the two is to be found in the “multiscale analysis” problems, which in the last half century have been the core of the studies in many areas of Analysis and in Physics, when theoretical developments and experimental techniques became finer and able to explore nature at smaller and smaller scales.

## Multiscale Analysis

To illustrate the multiscale analysis in PT it is convenient to present it in the context of Hamiltonian mechanics, because in this field it provides us with nontrivial cases of almost complete success.

We begin by contrasting the work of Siegel and that of Kolmogorov: which are based on radically different methods. The first being much closer in spirit to the developments of renormalization theory and to the Feynman graphs.

Most interesting formal PT series have a common feature: namely their  $n$ th order coefficients are constructed as sums of many “terms” and the first attempt to a complete analysis is to recognize that their sum, which gives the uniquely defined  $n$ th coefficient is much smaller than the sum of the absolute values of the constituent terms. This is a property usually referred to as a “cancellation” and, as a rule, it reflects some symmetry property of the problem: hence one possible approach is to look for expressions of the coefficients and for cancellations which would reduce the estimate of the  $n$ th order coefficients, very often of the order of a power of  $n!$ , to an exponential estimate  $O(q^{-n})$  for some  $q > 0$  yielding convergence (parenthetically in the mentioned case of Yang–Mills theories the reduction is even more dramatic as it leads from divergent expressions to finite ones, yet of order  $n!$ ).

The multiscale aspect becomes clear also in Kolmogorov’s method because the implicit functions theorem has to be applied over and over again and deals with functions implicitly defined on smaller and smaller domains (Gallavotti 1985a, 1986). But the method purportedly avoids

facing the combinatorial aspects behind the cancellations so much followed, and cherished ('t Hooft and Veltman 1972), in the Physics works.

Siegel's method was developed to study a problem in which no grouping of terms was eventually needed, even though this was by no means *a priori* (Pöschel 1986); and to realize that no cancellations were needed forced one to consider the problem as a multiscale one because the absence of rapid growth of the  $n$ th order coefficients became manifest after a suitable "hierarchical ordering" of the terms generating the coefficients. The approach establishes a strong connection with the Physics literature because the technique to study such cases was independently developed in quantum field theory with renormalization, as shown by Hepp in (Hepp 1966), relying strongly on it. This is very natural and, in case of failure, it can be improved by looking for "resummations" turning the power series into a convergent series in terms of functions of  $\varepsilon$  which are singular but controllably so. For details see below and (Gallavotti 1994).

What is "natural", however, is a very personal notion and it is not surprising that what some consider natural is considered unnatural or clumsy or difficult (or the three qualifications together) by others.

Conflict arises when the same problem can be solved by two different "natural" methods and in the case of PT for Hamiltonian systems close to integrable ones (closeness depending on the size of a parameter  $\varepsilon$ ), the so-called "small denominators" problem, the methods of Siegel and Kolmogorov are antithetic and an example of the just mentioned dualism.

The first method, that will be called here "Siegel's method" (see below for details), is based on a careful analysis of the structure of the various terms that occur at a given PT order achieving a proof that the  $n$ th order coefficient which is represented as the sum of many terms some of which *might* have size of order of a power of  $n!$  has in fact a size of  $O(q^{-n})$  so that the PT series is convergent for  $|\varepsilon| < q$ . Although strictly speaking the original work of Siegel does not immediately apply to the Hamiltonian

Mechanics problems (see below), it can nevertheless be adapted and yields a solution, as made manifest much later in (Eliasson 1996; Gallavotti 1994; Pöschel 1986).

The second method, called here "Kolmogorov's method", instead does not consider the individual coefficients of the various orders but just regards the sum of the series as a solution of an implicit function equation (a "Hamilton–Jacobi" equation) and devises a recursive algorithm approximating the unknown sum of the PT series by functions analytic in a disk of fixed radius  $q$  in the complex  $\varepsilon$ -plane (Gallavotti 1985a, 1986).

Of course the latter approach implies that no matter how we achieve the construction of the  $n$ th order PT series coefficient there will have to be enough cancellations, if at all needed, so that it turns out bounded by  $O(q^{-n})$ . And in the problem studied by Kolmogorov cancellations would be necessarily present if the  $n$ th order coefficient was represented by the sum of the terms in the Lindstedt series.

That this is not obvious is supported by the fact that it was considered an open problem, for about 30 years, to find a way to exhibit explicitly the cancellation mechanism in the Lindstedt series implied by Kolmogorov's work. This was done by Eliasson (Eliasson 1996), who proved that the coefficients of the PT of a given order  $n$  as expressed by the construction known as the "Lindstedt algorithm" yielded coefficients of size of  $O(q^{-n})$ : his argument, however, did not identify in general which term of the Lindstedt sum for the  $n$ th order coefficient was compensated by which other term or terms. It proved that the sum had to satisfy suitable relations, which in turn implied a total size of  $O(q^{-n})$ . And it took a few more years for the complete identification (Gallavotti 1994), of the rules to follow in collecting the terms of the Lindstedt series which would imply the needed cancellations.

It is interesting to remark that, aside from the example of Hamiltonian PT, multiscale problems have dominated the development of analysis and Physics in recent time: for instance they appear in harmonic analysis (Carleson, Fefferman), in PDE's (DeGiorgi, Moser, Caffarelli–Kohn–Nirenberg), in relativistic quantum mechanics (Glimm, Jaffe, Wilson), in Hamiltonian Mechanics (Siegel,

Kolmogorov, Arnold, Moser), in statistical mechanics and condensed matter (Fisher, Wilson, Widom) . . . Sometimes, although not always, studied by PT techniques (Gallavotti 2001).

## A Paradigmatic Example of PT Problem

It is useful to keep in mind an example illustrating technically what it means to perform a multiscale analysis in PT. And the case of quasiperiodic motions in Hamiltonian mechanics will be selected here, being perhaps the simplest.

Consider the motion of  $\ell$  unit masses on a unit circle and let  $\alpha = (\alpha_1, \dots, \alpha_\ell)$  be their positions on the circle, i.e.,  $\alpha$  is a point on the torus  $\mathcal{T}^\ell = [0, 2\pi]^\ell$ . The points interact with a potential energy  $\varepsilon f(\alpha)$  where  $\varepsilon$  is a strength parameter and  $f$  is a trigonometric even polynomial, of degree  $N : f(\alpha) = \sum_{\mathbf{v} \in \mathbb{Z}^\ell, |\mathbf{v}| \leq N} f_{\mathbf{v}} e^{i\mathbf{v} \cdot \alpha}$ ,  $f_{\mathbf{v}} = f_{-\mathbf{v}} \in \mathbb{R}$ , where  $\mathbb{Z}^\ell$  denotes the lattice of the points with integer components in  $\mathbb{R}^\ell$  and  $|\mathbf{v}| = \sum_j |v_j|$ .

Let  $t \rightarrow \alpha_0 + \omega_0 t$  be the motion with initial data, at time  $t = 0$ ,  $\alpha(0) = \alpha_0$ ,  $\dot{\alpha}(0) = \omega_0$ , in which all particles rotate at constant speed with rotation velocity  $\omega_0 = (\omega_{01}, \dots, \omega_{0\ell}) \in \mathbb{R}^\ell$ . This is a solution for the equations of motion for  $\varepsilon = 0$  and it is a quasi periodic solution, i.e., each of the angles  $\alpha_j$  rotates periodically at constant speed  $\omega_{0j}$ ,  $j = 1, \dots, \ell$ .

The motion will be called non resonant if the components of the rotation speed  $\omega_0$  are rationally independent: this means that  $\omega_0 \cdot \mathbf{v} = 0$  with  $\mathbf{v} \in \mathbb{Z}^\ell$  is possible only if  $\mathbf{v} = \mathbf{0}$ . In this case the motion  $t \rightarrow \alpha_0 + \omega_0 t$  covers,  $\forall \alpha_0$ , densely the torus  $\mathcal{T}^\ell$  as  $t$  varies. The PT problem that we consider is to find whether there is a family of motions “of the same kind” for each  $\varepsilon$ , small enough, solving the equations of motion; more precisely whether there exists a function  $\mathbf{a}_\varepsilon(\varphi)$ ,  $\varphi \in \mathcal{T}^\ell$ , such that setting

$$\alpha(t) = \varphi + \omega_0 t + \mathbf{a}_\varepsilon(\varphi + \omega_0 t), \quad \text{for } \varphi \in \mathcal{T}^\ell \quad (1)$$

one obtains,  $\forall \varphi \in \mathcal{T}^\ell$  and for  $\varepsilon$  small enough, a solution of the equations of motion for a force  $-\varepsilon \partial_\alpha f(\alpha)$ : i.e.,

$$\ddot{\alpha}(t) = -\varepsilon \partial_\alpha f(\alpha(t)). \quad (2)$$

By substitution of Eq. (1) in Eq. (2), the condition becomes  $(\omega_0 \cdot \partial_\varphi)^2 \mathbf{a}(\varphi + \omega_0 t) = -\partial_\alpha f(\varphi + \omega_0 t)$ . Since  $\omega_0$  is assumed rationally independent  $\varphi + \omega_0 t$  covers densely the torus  $\mathcal{T}^\ell$  as  $t$  varies: hence the equation for  $\mathbf{a}_\varepsilon$  is

$$(\omega_0 \cdot \partial_\varphi)^2 \mathbf{a}_\varepsilon(\varphi) = -\varepsilon \partial_\alpha f(\varphi + \mathbf{a}_\varepsilon(\varphi)) \quad (3)$$

Applying PT to this equation means to look for a solution  $\mathbf{a}_\varepsilon$  which is analytic in  $\varepsilon$  small enough and in  $\varphi \in \mathcal{T}^\ell$ . In colorful language one says that the perturbation effect is slightly deforming a non-resonant torus with given frequency spectrum (i.e. given  $\omega_0$ ) on which the motion develops, without destroying it and keeping the quasi periodic motion on it with the same frequency spectrum.

## Lindstedt Series

As it follows from a very simple special case of Poincaré’s work, Eq. (3) cannot be solved if also  $\omega_0$  is considered variable and the dependence on  $\varepsilon$ ,  $\omega_0$  analytic. Nevertheless if  $\omega_0$  is fixed and non resonant and if  $\mathbf{a}_\varepsilon$  is supposed analytic in  $\varepsilon$  small enough and in  $\varphi \in \mathcal{T}^\ell$ , then there can be at most one solution to the Eq. (3) with  $\mathbf{a}_\varepsilon(\mathbf{0}) = \mathbf{0}$  (which is not a real restriction because if  $\mathbf{a}_\varepsilon(\varphi)$  is a solution also  $\mathbf{a}_\varepsilon(\varphi + \mathbf{c}_\varepsilon) + \mathbf{c}_\varepsilon$  is a solution for any constant  $\mathbf{c}_\varepsilon$ ). This so because the coefficients of the power series in  $\varepsilon$ ,  $\sum_{n=1}^{\infty} \varepsilon^n \mathbf{a}_n(\varphi)$ , are uniquely determined if the series is convergent. In fact they are trigonometric polynomials of order  $\leq nN$  which will be written as

$$\alpha_n(\varphi) = \sum_{0 < |\mathbf{v}| \leq Nn} \alpha_{n,\mathbf{v}} e^{i\mathbf{v} \cdot \varphi}. \quad (4)$$

It is convenient to express them in terms of graphs. The graphs to use to express the value  $\alpha_{n,\mathbf{v}}$  are

- (i) trees with  $n$  nodes  $v_1, \dots, v_n$ ,
- (ii) one root  $r$ ,
- (iii)  $n$  lines joining pairs  $(v'_i, v_i)$  of nodes or the root and one node, *always one and not more*,  $(r, v_i)$ ;

- (iv) the lines will be different from each other and distinguished by a mark label,  $1, \dots, n$  attached to them. The connections between the nodes that the lines generate have to be loopless, i.e., the graph formed by the lines must be a tree.
- (v) The tree lines will be imagined oriented towards the root: hence a partial order is generated on the tree and the line joining  $v$  to  $v'$  will be denoted  $\lambda_{v,v'}$  and  $v'$  will be the node closer to the root immediately following  $v$ , hence such that  $v' > v$  in the partial order of tree.

The number of such trees is large and exactly equal to  $n^{n-1}$ , as an application of Cayley's formula implies: their collection will be denoted  $T_n^0$ .

To compute  $\mathbf{a}_{n,\mathbf{v}}$  consider all trees in  $T_n^0$  and attach to each node  $v$  a vector  $\mathbf{v}_v \in \mathbb{Z}^\ell$ , called "mode label", such that  $f_v \neq 0$ , hence  $|\mathbf{v}_v| \leq N$ . To the root we associate one of the coordinate unit vectors  $\mathbf{v}_r \equiv \mathbf{e}_r$ . We obtain a set  $T_n$  of *decorated trees* (with  $\leq (2N+1)^{\ell n} n^{n-1}$  elements, by the above counting analysis).

Given  $\theta \in T_n$  and  $\lambda = \lambda(v', v) \in \theta$  we define the *current* on the line  $\lambda$  to be the vector  $\mathbf{v}(\lambda) \equiv \mathbf{v}(v', v) \stackrel{\text{def}}{=} \sum_{w \leq v} \mathbf{v}_w$ : i.e., we imagine that the node vectors  $\mathbf{v}_{v_i}$  represent currents entering the node  $v_i$  and flowing towards the root. Then  $\mathbf{v}(\lambda)$  is, for each  $\lambda$ , the sum of the currents which entered all the nodes not following  $v$ , i.e., current accumulated after passing the node  $v$ .

The current flowing in the root line  $v = \sum_v \mathbf{v}_v$  will be denoted  $\mathbf{v}(\theta)$ .

Let  $T_n^*$  be the set trees in  $T_n$  in which *all lines* carry a *non zero* current  $\mathbf{v}(\lambda) \neq \mathbf{0}$ . A value  $\text{Val}(\theta)$  will be defined, for  $\theta \in T_n^*$ , by a product of node factors and of line factors over all nodes and

$$\text{Val}(\theta) = \frac{i(-1)^n}{n!} \prod_{v \in \theta} f_{\mathbf{v}_v} \prod_{\lambda=(v', v)} \frac{\mathbf{v}_{v'} \cdot \mathbf{v}_v}{(\boldsymbol{\omega}_0 \cdot \mathbf{v}(v', v))^2} \quad (5)$$

The coefficient  $\mathbf{a}_{n,\mathbf{v}}$  will then be

$$\mathbf{a}_{n,\mathbf{v}} = \sum_{\substack{\theta \in T_n^* \\ \mathbf{v}(\theta) = \mathbf{v}}} \text{Val}(\theta) \quad (6)$$

and, when the coefficients are imagined to be constructed in this way, the formal power series  $\sum_{n=1}^{\infty} \varepsilon^n \sum_{|\mathbf{v}| \leq Nn} \mathbf{a}_{n,\mathbf{v}}$  is called the "Lindstedt series". Eq. (5) and its graphical interpretation in Fig. 1 should be considered the "Feynman rules" and the "Feynman diagrams" of the PT for Eq. (3) (Gallavotti 1995, 2001).

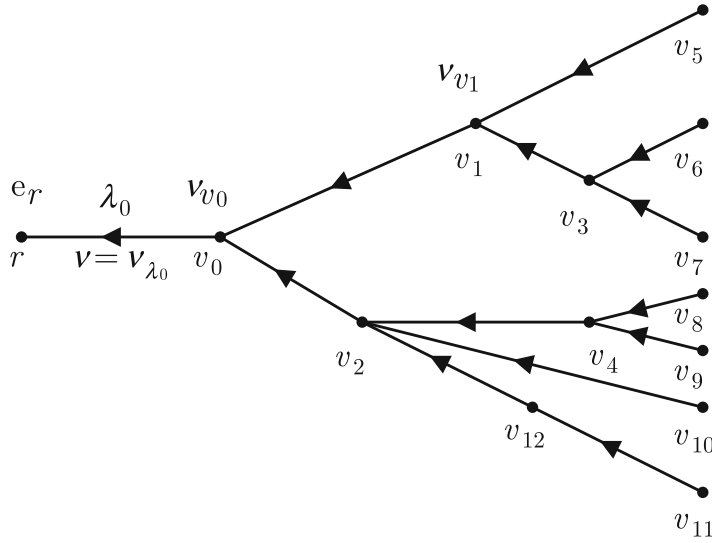
## Convergence. Scales. Multiscale Analysis

The Lindstedt series is well defined because of the non resonance condition and the  $n$ th term is not even a sum of too many terms: if  $F \stackrel{\text{def}}{=} \max_{\mathbf{v}} |f_{\mathbf{v}}|$ , each of them can be bounded by  $F^n / n! \prod_{\lambda \in \theta} N^2 / (\boldsymbol{\omega}_0 \cdot \mathbf{v}(\lambda)^2)$ ; hence their sum can be bounded, if  $G$  is such that  $((2N+1)^{\ell n} n^{n-1} F^n) / n! \leq G^n$ , by  $G^n \prod_{\lambda \in \theta} N^2 / (\boldsymbol{\omega}_0 \cdot \mathbf{v}(\lambda)^2)$ .

Thus, all  $\mathbf{a}_n$  are well defined and finite *but* the problem is that  $|\mathbf{v}(\lambda)|$  can be large (up to  $Nn$  at given order  $n$ ) and therefore  $\boldsymbol{\omega}_0 \cdot \mathbf{v}(\lambda)$  although never zero can become very small as  $n$  grows. For this reason the problem of convergence of the series is an example of what is called a *small denominators problem*. And it is necessary to assume more than just non resonance of  $\boldsymbol{\omega}_0$  in order to solve it in the present case: a simple condition is the Diophantine condition, namely the existence of  $C, \tau > 0$  such that

$$|\boldsymbol{\omega}_0 \cdot \mathbf{v}| \geq \frac{1}{C|\mathbf{v}|^\tau}, \quad \forall \mathbf{0} \neq \mathbf{v} \in \mathbb{Z}^\ell. \quad (7)$$

But this condition is not sufficient in an obvious way: because it only allows us to bound individual tree-values by  $n!^a$  for some  $a > 0$  related to  $\tau$ ; furthermore it is not difficult to check that there are single graphs whose value is actually of "factorial" size in  $n$ . Although non trivial to see (as mentioned above) this was only apparently so in the earlier case of Siegel's



**Perturbation Theory, Fig. 1** A tree  $\theta$  with  $m_{v_0}=2, m_{v_1}=2, m_{v_2}=3, m_{v_3}=2, m_{v_4}=2, m_{v_{12}}=1$  lines entering the nodes  $v_i, k=13$ . Some labels or decorations explicitly marked (on the lines  $\lambda_0, \lambda_1$  and on the nodes  $v_1, v_2$ ); the

number labels, distinguishing the branches, are not shown. The arrows represent the partial ordering on the tree

problem but it is the new essential feature of the terms generating the  $n$ th order coefficient in Eq. (6).

A resummation is necessary to show that the tree-values can be grouped so that the sum of the values of each group can be bounded by  $q^{-n}$  for some  $q > 0$  and  $\forall n$ , although the group may contain (several) terms of factorial size. The terms to be grouped have to be ordered hierarchically according to the sizes of the line factors  $1/(\omega_0 \cdot \nu(\lambda))^2$ , which are called *propagators* in (Gallavotti 1994; Gallavotti et al. 2004).

A similar problem is met in quantum field theory where the graphs are the Feynman graphs: such graphs can only have a small number of lines that converge into a node but they can have loops, and to show that the perturbation series is well defined to all orders it is also necessary to collect terms hierarchically according to the propagators sizes. The systematic way was developed by Hepp (Hepp 1966, 1969), for the PT expansion of the Schwinger functions in quantum field theory of scalar fields (Gallavotti 1985b). It has been used on many occasions later and it plays a key role in the renormalization

group methods in Statistical Mechanics (for instance in theory of the ground state of Fermi systems) (Gallavotti 2001; Gallavotti and Benfatto 1995).

However, it is in the Lindstedt series that the method is perhaps best illustrated. Essentially because it ends up in a convergence proof, while often in the field theory or statistical mechanics problems the PT series can be only proved to be well defined to all orders, but they are seldom, if ever, convergent so that one has to have recourse to other supplementary analytic means to show that the PT series are asymptotic (in the cases in which they are such).

The path of the proof is the following.

- (1) Consider only trees in which no two lines  $\lambda_+$  and  $\lambda_-$ , with  $\lambda_+$  following  $\lambda_-$  in the partial order of the tree, have the *same* current  $\nu_0$ . In this case the maximum of the  $\prod_{\lambda} 1/(\omega_0 \cdot \nu(\lambda))^2$  over all trees  $\theta \in T_n^*$  can be bounded by  $G_1^n$  for some  $G_1$ .

This is an immediate consequence and the main result in Siegel's original work (Siegel



1943), which dealt with a different problem with small denominators in its formal PT solution: the coefficients of the series could also be represented by tree graphs, very similar too the ones above: but the only allowed  $\mathbf{v} \in \mathbb{Z}^\ell$  were the non zero vectors with all components  $\geq 0$ .

The latter property automatically guarantees that the graphs contain no pair of lines  $\lambda_+$ ,  $\lambda_-$  following each other as above in the tree partial order and having the same current. Siegel's proof also implies a multiscale analysis (Pöschel 1986): but it requires no grouping of the terms unlike the analogue Lindstedt series, Eq. (6).

- (2) Trees which contain lines  $\lambda_+$  and  $\lambda_-$ , with  $\lambda_+$  following  $\lambda_-$ , in the partial order of the tree, and having the *same* current  $\mathbf{v}_0$  can have values which have size of order  $O(n!^a)$  with some  $a > 0$ . Collecting terms is the refore essential.

A line  $\lambda$  of a tree is said to have scale  $k$  if  $2^{-k-1} \leq 1/C |\boldsymbol{\omega}_0 \cdot \mathbf{v}| < 2^{-k}$ . The lines of a tree  $\theta \in T_n^*$  can then be collected in clusters (The scaling factor 2 is arbitrary: any scale factor  $> 1$  could be used).

A cluster of scale  $p$  is a maximal connected set of lines of scale  $k \geq p$  with at least one line of scale  $p$ . Clusters are connected to the rest of the tree by lines of lower scale which can be *incoming* or *outgoing* with respect to the partial ordering. Clusters also contain nodes: a node is in a cluster if it is an extreme of a line contained in a cluster; such nodes are said to be *internal* to the cluster.

Of particular interest are the self energy clusters. These are clusters with only one incoming line and only one outgoing line which *furthermore* have the same current  $\mathbf{v}_0$ .

To simplify the analysis the Diophantine condition can be strengthened to insure that if in a tree graph the line incoming into a self energy cluster and ending in an internal node  $\mathbf{v}$  is detached from the node  $\mathbf{v}$  and reattached to another node internal to the same cluster which is not in a self-energy subcluster (if any) then the new tree nodes are still enclosed in the same clusters. Alternatively the definition of scale of a line can be modified slightly to achieve the same goal.

- (3) Then it makes sense to sum together all the values of the trees whose nodes are collected into the same families of clusters and differ only because the lines entering the self energy clusters are attached to a different node internal to the cluster, but external to the inner self energy subclusters (if any). Furthermore, the value of the trees obtained by changing simultaneously sign to the  $\mathbf{v}_v$  of the nodes inside the self energy clusters have also to be added together.

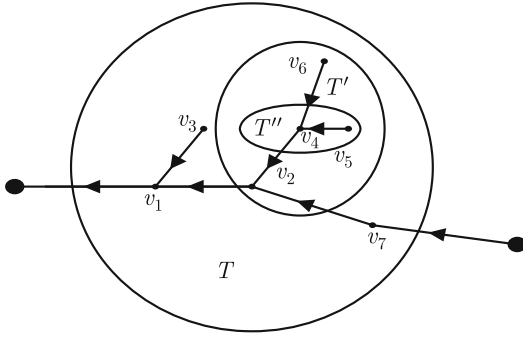
After collecting the terms in the described way it is possible to check that each sum of terms so collected is bounded by  $Q_0^{-n}$  for some  $Q_0$  (which can also be estimated explicitly). Since the number of addends left is not larger than the original one the bound on  $\sum |\mathbf{a}_{n,\mathbf{v}}|$  becomes  $\leq \left( F^n (2N+1)^\ell n^n N^{2n} \right) / n! Q_0^{-n} \leq Q^{-n}$ , for suitable  $Q_0, Q$ , so that convergence of the formal series for  $\mathbf{a}_\varepsilon(\boldsymbol{\varphi})$  is achieved for  $|\varepsilon| < Q$ , see (Gallavotti 1994).

## Non Convergent Cases

Convergence is *not* the rule: very interesting problems arise in which the PT series is, or is believed to be, only asymptotic. For instance in quantum field theory the PT series are well defined but they are not convergent: they can be proved, in the scalar  $\varphi^4$  theories in dimension 2 and 3 to be asymptotic series for a function of  $\varepsilon$  which is Borel summable: this means in particular that the solution can be in principle recovered, for  $\varepsilon > 0$  and small, just from the coefficients of its formal expansion (Fig. 2).

Other non convergent expansions occur in statistical mechanics, for example in the theory of the ground state of a Fermi gas of particles on a lattice of obstacles. This is still an open problem, and a rather important one. Or occur in quantum field theory where sometimes they can be proved to be Borel summable.

The simplest instances again arise in Mechanics in studying resonant quasi periodic motions. A paradigmatic case is provided by Eqs. (1) and (2) when  $\boldsymbol{\omega}_0$  has some vanishing components:



**Perturbation Theory, Fig. 2** An example of three clusters symbolically delimited by circles, as visual aids, inside a tree (whose remaining branches and clusters are not drawn and are indicated by the bullets); not all labels are explicitly shown. The scales (not marked) of the branches increase as one crosses inward the circles boundaries: recall, however, that the scale labels are integers  $\leq 1$  (hence typically  $\leq 0$ ). The  $\nu$  labels are not drawn (but must be imagined). If the  $\nu$  labels of  $(\nu_4, \nu_5)$  add up to 0 the cluster  $T''$  is a self-energy graph. If the  $\nu$  labels of  $(\nu_2, \nu_4, \nu_5, \nu_6)$  add up to 0 the cluster  $T'$  is a self-energy graph and such is  $T$  if the  $\nu$  labels of  $(\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7)$  add up to 0. The cluster  $T'$  is maximal in  $T$

$\omega_0 = (\omega_1, \dots, \omega_r, 0, \dots, 0) = (\tilde{\omega}_0, 0)$  with  $1 < r < \ell$ . If one writes  $\alpha = (\tilde{\alpha}, \tilde{\beta}) \in \mathcal{T}^r \times \mathcal{T}^{\ell-r}$  and looks motions like Eq. (1) of the form

$$\begin{aligned} \tilde{\alpha}(t) &= \tilde{\varphi} + \tilde{\omega}_0 t + \tilde{\mathbf{a}}_\varepsilon(\tilde{\varphi} + \tilde{\omega}_0 t) \\ \tilde{\beta}(t) &= \beta_0 + \tilde{\mathbf{b}}_\varepsilon(\tilde{\varphi} + \tilde{\omega}_0 t) \end{aligned} \quad (8)$$

where  $\tilde{\mathbf{a}}_\varepsilon(\tilde{\varphi}), \tilde{\mathbf{b}}_\varepsilon(\tilde{\varphi})$  are functions of  $\tilde{\varphi} \in \mathcal{T}^r$ , analytic in  $\varepsilon$  and  $\tilde{\varphi}$ .

In this case the analogue of the Lindstedt series can be devised provided  $\beta_0$  is chosen to be a stationary point for the function  $\tilde{f}(\tilde{\beta}) = \int f(\tilde{\alpha}, \tilde{\beta}) \frac{d\tilde{\alpha}}{(2\pi)^r}$ , and provided  $\tilde{\omega}_0$  satisfies a Diophantine property  $|\tilde{\omega}_0 \cdot \tilde{\mathbf{v}}| > 1/(C|\tilde{\mathbf{v}}|^\tau)$  for all  $0 \neq \tilde{\mathbf{v}} \in \mathbb{Z}^r$  and for  $\tau, C$  suitably chosen.

This time the series is likely to be, in general, non convergent (although there is not a proof yet). And the terms of the Lindstedt series can be suitably collected to improve the estimates. Nevertheless, the estimates cannot be improved enough to obtain convergence. Deeper resummations are needed to show that in some cases the terms of the series can be collected and rearranged into a convergent series.

The resummation is deeper in the sense that it is not enough to collect terms contributing to a given order in  $\varepsilon$  but it is necessary to collect and sum terms of different order according to the following scheme.

1. The terms of the Lindstedt series are first “regularized” so that the new series is manifestly analytic in  $\varepsilon$  with, however, a radius of convergence depending on the regularization. For instance one can consider only terms with lines of scale  $\leq M$ .
2. Terms of different orders in  $\varepsilon$  are then summed together and the series becomes a series in powers of functions  $\lambda_j(\varepsilon; M)$  of  $\varepsilon$  with very small radius of convergence in  $\varepsilon$ , but with an  $M$ -independent radius of convergence  $\varrho$  in the  $\lambda_j(\varepsilon, M)$ . The labels  $j = 0, 1, \dots, M$  are scale labels whose value is determined by the order in which they are generated in the hierarchical organization of the collection of the graphs according to their scales.
3. One shows that the functions  $\lambda_j(\varepsilon; M)$  (“running couplings”) can be analytically continued in  $\varepsilon$  to an  $M$ -independent domain  $\mathcal{D}$  containing the origin in its closure and where they remain smaller than  $\varrho$  for all  $M$ . Furthermore,  $\lambda_j(\varepsilon; M) \xrightarrow{M \rightarrow \infty} \lambda_j(\varepsilon)$ , for  $\varepsilon \in \mathcal{D}$ .
4. The convergent power series in the running couplings admits an asymptotic series in  $\varepsilon$  at the origin which coincides with the formal Lindstedt series. Hence in the domain  $\mathcal{D}$  a meaning is attributed to the sum of Lindstedt series.
5. One checks that the functions  $\tilde{\mathbf{a}}_\varepsilon, \tilde{\mathbf{b}}_\varepsilon$  thus defined are such that Eq. (8) satisfies the equations of motion Eq. (2).

The proof can be completed if the domain  $\mathcal{D}$  contains real points  $\varepsilon$ .

If  $\tilde{\beta}_0$  is a maximum point the domain  $\mathcal{D}$  contains a circle tangent to the origin and centered on the positive real axis. So in this case the  $\tilde{\mathbf{a}}_\varepsilon, \tilde{\mathbf{b}}_\varepsilon$  are constructed in  $\mathcal{D} \cap \mathcal{R}_+, \mathcal{R}_+ \stackrel{\text{def}}{=} (0, +\infty)$ .

If instead  $\tilde{\beta}_0$  is a minimum point the domain  $\mathcal{D}$  exists but  $\mathcal{D} \cap \mathcal{R}_+$  touches the positive real axis on a set of points with positive measure and density 1 at the origin. So  $\tilde{\mathbf{a}}_\varepsilon, \tilde{\mathbf{b}}_\varepsilon$  are constructed only



for  $\varepsilon$  in this set which is a kind of “Cantor set” (Gallavotti and Gentile 2005).

Again the multiscale analysis is necessary to identify the tree values which have to be collected to define  $\lambda_j(\varepsilon; M)$ . In this case it is an analysis which is much closer to the similar analysis that is encountered in quantum field theory in the “self energy resummations”, which involve collecting and summing graph values of graphs contributing to different orders of perturbation.

The above scheme can also be applied when  $r = \ell$ , i.e., in the case of the classical Lindstedt series when it is actually convergent: this leads to an alternative proof of the Kolmogorov theorem which is interesting as it is even closer to the renormalization group methods because it expresses the solution in terms of a power series in running couplings (Chaps. 8, 9 in 12).

## Conclusion and Outlook

Perturbation theory provides a general approach to the solution of problems “close” to well understood ones, “closeness” being measured by the size of a parameter  $\varepsilon$ . It naturally consists of two steps: the first is to find a formal solution, under the assumption that the quantities of interest are analytic in  $\varepsilon$  at  $\varepsilon = 0$ . If this results in a power series with well-defined coefficients then it becomes necessary to find whether the series thus constructed, called a formal series, converges.

In general the proof that the formal series exists (when it really does) is nontrivial: typically in quantum mechanics problems (quantum fields or statistical mechanics) this is an interesting and deep problem giving rise to renormalization theory. Even in classical mechanics PT of integrable systems it has been, historically, a problem to obtain (in wide generality) the Lindstedt series (of which a simple example is discussed above).

Once existence of a PT series is established, very often the series is not convergent and at best is an asymptotic series. It becomes challenging to find its meaning (if any, as there are cases, even interesting ones, on which conjectures exist claiming that the series have no meaning, like

the quantum scalar field in dimension 4 with “ $\varphi^4$ -interaction” or quantum electrodynamics).

Convergence proofs, in most interesting cases, require a multiscale analysis: because the difficulty arises as a consequence of the behavior of singularities at infinitely many scales, as in the case of the Lindstedt series above exemplified.

When convergence is not possible to prove, the multiscale analysis often suggests “resummations”, collecting the various terms whose sum yields the formal PT series (usually the algorithms generating the PT series give its terms at given order as sums of simple but many quantities, as in the discussed case of the Lindstedt series). The collection involves adding together terms of different order in  $\varepsilon$  and results in a new power series, the resummed series, in a family of parameters  $\lambda_j(\varepsilon)$  which are functions of  $\varepsilon$ , called the “running couplings”, depending on a “scale index”  $j = 0, 1, \dots$

The running couplings are (in general) singular at  $\varepsilon = 0$  as functions of  $\varepsilon$  but  $C^\infty$  there, and obey equations that allow one to study and define them independently of a convergence proof. If the running couplings can be shown to be so small, as  $\varepsilon$  varies in a suitable domain  $\mathcal{D}$  near 0, to guarantee convergence of the resummed series and therefore to give a meaning to the PT for  $\varepsilon \in \mathcal{D}$  then the PT program can be completed.

The singularities in  $\varepsilon$  at  $\varepsilon = 0$  are therefore all contained in the running couplings, usually very few and the same for various formal series of interest in a given problem.

The idea of expressing the sum of formal series as sum of convergent series in new parameters, the running couplings, determined by other means (a recursion relation denominated the beta function flow) is the key idea of the renormalization group methods: PT in mechanics is a typical and simple example.

On purpose attention has been devoted to PT in the analytic class: but it is possible to use PT techniques in problems in which the functions whose value is studied are not analytic; the techniques are somewhat different and new ideas are needed which would lead quite far away from the natural PT framework which is within the analytic class.

Of course there are many problems of PT in which the formal series are simply convergent and the proof does not require any multiscale analysis. Greater attention was devoted to the novel aspect of PT that emerged in Physics and Mathematics in the last half century and therefore problems not requiring multiscale analysis were not considered. It is worth mentioning, however, that even in simple convergent PT cases it might be convenient to perform resummations. An example is Kepler's equation

$$\ell = \xi - \varepsilon \sin \xi, \quad \xi, \ell \in T^1 = [0, 2\pi] \quad (9)$$

which can be (easily) solved by PT. The resulting series has a radius of convergence in  $\varepsilon$  rather small (Laplace's limit): however if a resummation of the series is performed transforming it into a power series in a "running coupling"  $\lambda_0(\varepsilon)$  (only 1, because no multiscale analysis is needed, the PT series being convergent) given by (Vol. 2, p. 321 in (Levi-Civita 1956))

$$\lambda_0 \stackrel{\text{def}}{=} \frac{\varepsilon e^{\sqrt{1-\varepsilon^2}}}{1 + \sqrt{1-\varepsilon^2}}, \quad (10)$$

then the resummed series is a power series in  $\lambda_0$  with radius of convergence 1 and when  $\varepsilon$  varies between 0 and 1 the parameter  $\lambda_0$  corresponding to it goes from 0 to 1. Hence in terms of  $\lambda_0$  it is possible to invert *by power series* the Kepler equation for all  $\varepsilon \in [0, 1)$ , i.e., in the entire interval of physical interest (recall that  $\varepsilon$  has the interpretation of eccentricity of an elliptic orbit in the 2-body problem). Resummations can improve convergence properties.

## Future Directions

It is always hard to indicate future directions, which usually turn to different paths. Perturbation theory is an ever evolving subject: it is a continuous source of problems and its applications generate new ones. Examples of outstanding problems are understanding the triviality conjectures of models like quantum  $\varphi^4$  field theory in dimension 4 (Gallavotti 1985b); or a development of the theory of the ground states

of Fermionic systems in dimensions 2 and 3 (Gallavotti and Benfatto 1995); a theory of weakly coupled Anosov flows to obtain information of the kind that it is possible to obtain for weakly coupled Anosov maps (Gallavotti et al. 2004); uniqueness issues in cases in which PT series can be given a meaning, but in a priori non unique way like the resonant quasi periodic motions in nearly integrable Hamiltonian systems (Gallavotti et al. 2004).

## Bibliography

- Einstein A (1917) Zum Quantensatz von Sommerfeld und Epstein. *Verh Dtsch Phys Ges* 19:82–102
- Eliasson LH (1996) Absolutely convergent series expansions for quasi-periodic motions. *Math Phys Electron J (MPEJ)* 2:33
- Fermi E (1923) Il principio delle adiabatiche e i sistemi che non ammettono coordinate angolari. *Nuovo Cimento* 25:171–175. Reprinted in *Collected papers*, vol I, pp 88–91
- Fröhlich J (1982) On the triviality of  $\lambda\varphi_d^4$  theories and the approach to the critical point in  $d(\geq 4)$  dimensions. *Nucl Phys B* 200:281–296
- Gallavotti G (1985a) Perturbation theory for classical Hamiltonian systems. In: Fröhlich J (ed) *Scaling and self similarity in physics*. Birkhauser, Boston
- Gallavotti G (1985b) Renormalization theory and ultraviolet stability for scalar fields via renormalization group methods. *Rev Mod Phys* 57:471–562
- Gallavotti G (1986) Quasi integrable mechanical systems. In: *Phénomènes Critiques, Systèmes aléatoires, Théories de jauge* (Proceedings, Les Houches, XLIII (1984)), vol II. North Holland, Amsterdam, pp 539–624
- Gallavotti G (1994) Twistless KAM tori. *Commun Math Phys* 164:145–156
- Gallavotti G (1995) Invariant tori: a field theoretic point of view on Eliasson's work. In: Figari R (ed) *Advances in dynamical systems and quantum physics*. World Scientific, pp 117–132
- Gallavotti G (2001) Renormalization group in statistical mechanics and mechanics: gauge symmetries and vanishing beta functions. *Phys Rep* 352:251–272
- Gallavotti G, Benfatto G (1995) *Renormalization group*. Princeton University Press, Princeton
- Gallavotti G, Gentile G (2005) Degenerate elliptic resonances. *Commun Math Phys* 257:319–362. <https://doi.org/10.1007/s00220-005-1325-6>
- Gallavotti G, Bonetto F, Gentile G (2004) *Aspects of the ergodic, qualitative and statistical theory of motion*. Springer, Berlin
- Glimm J, Jaffe A (1981) *Quantum physics. A functional integral point of view*. Springer, New York
- Gross DJ (1999) Twenty five years of asymptotic freedom. *Nucl Phys B (Proc Suppl)* 74:426–446. [https://doi.org/10.1016/S0920-5632\(99\)00208-X](https://doi.org/10.1016/S0920-5632(99)00208-X)

- Hepp K (1966) Proof of the Bogoliubov–Parasiuk theorem on renormalization. *Commun Math Phys* 2:301–326
- Hepp K (1969) *Théorie de la renormalization*. Lecture notes in Physics, vol 2. Springer, Berlin
- ‘t Hooft G (1999) When was asymptotic freedom discovered? or The rehabilitation of quantum field theory. *Nucl Phys B (Proc Suppl)* 74:413–425. [https://doi.org/10.1016/S0920-5632\(99\)00207-8](https://doi.org/10.1016/S0920-5632(99)00207-8)
- ‘t Hooft G, Veltman MJG (1972) Regularization and renormalization of gauge fields. *Nucl Phys B* 44: 189–213
- Kolmogorov AN (1954) On the preservation of conditionally periodic motions. *Dokl Akad Nauk SSSR* 96: 527–530. and in: Casati G, Ford J (eds) (1979) *Stochastic behavior in classical and quantum Hamiltonians*. Lecture Notes in Physics vol 93. Springer, Berlin
- Laplace PS (1799) *Mécanique Céleste*. Paris. Reprinted by Chelsea, New York
- Levi-Civita T (1956) *Opere Matematiche*, vol 2. Accademia Nazionale dei Lincei and Zanichelli, Bologna
- Morrey CB (1955) On the derivation of the equations of hydrodynamics from statistical mechanics. *Commun Pure Appl Math* 8:279–326
- Nelson E (1966) A quartic interaction in two dimensions. In: Goodman R, Segal I (eds) *Mathematical theory of elementary particles*. MIT, Cambridge, pp 69–73
- Poincaré H (1892) *Les Méthodes nouvelles de la Mécanique céleste*. Paris. Reprinted by Blanchard, Paris, 1987
- Pöschel J (1986) Invariant manifolds of complex analytic mappings. In: Osterwalder K, Stora R (eds) *Phénomènes Critiques, Systèmes aleatoires, Théories de jauge* (Proceedings, Les Houches, XLIII (1984)), vol II. North Holland, Amsterdam, pp 949–964
- Siegel K (1943) Iterations of analytic functions. *Ann Math* 43:607–612
- Simon B (1974) *The  $P(\varphi)_2$  Euclidean (quantum) field theory*. Princeton University Press, Princeton
- Wilson K, Kogut J (1973) The renormalization group and the  $\varepsilon$ -expansion. *Phys Rep* 12:75–199

# Hamiltonian Perturbation Theory (and Transition to Chaos)

Henk W. Broer<sup>1</sup> and Heinz Hanßmann<sup>2</sup>

<sup>1</sup>Instituut voor Wiskunde en Informatica, Rijksuniversiteit Groningen, Groningen, The Netherlands

<sup>2</sup>Mathematisch Instituut, Universiteit Utrecht, Utrecht, The Netherlands

## Article Outline

Glossary

Definition of the Subject

Introduction

One Degree of Freedom

Perturbations of Periodic Orbits

Invariant Curves of Planar Diffeomorphisms

KAM Theory: An Overview

Splitting of Separatrices

Transition to Chaos and Turbulence

Future Directions

Bibliography

## Glossary

**Bifurcation** In parametrized dynamical systems a bifurcation occurs when a qualitative change is invoked by a change of parameters. In models such a qualitative change corresponds to transition between dynamical regimes. In the generic theory a finite list of cases is obtained, containing elements like ‘saddle-node’, ‘period doubling’, ‘Hopf bifurcation’ and many others.

**Cantor set, Cantor dust, Cantor family, Cantor stratification** Cantor dust is a separable locally compact space that is perfect, i.e. every point is in the closure of its complement, and totally disconnected. This determines Cantor dust up to homeomorphisms. The term Cantor set (originally reserved for the specific form of

Cantor dust obtained by repeatedly deleting the middle third from a closed interval) designates topological spaces that locally have the structure  $\mathbb{R}^n \times \text{Cantor dust}$  for some  $n \in \mathbb{N}$ . Cantor families are parametrized by such Cantor sets. On the real line  $\mathbb{R}$  one can define Cantor dust of positive measure by excluding around each rational number  $p/q$  an interval of size

$$\frac{2\gamma}{q^\tau}, \quad \gamma > 0, \tau > 2.$$

Similar Diophantine conditions define Cantor sets in  $\mathbb{R}^n$ . Since these Cantor sets have positive measure their Hausdorff dimension is  $n$ . Where the unperturbed system is stratified according to the co-dimension of occurring (bifurcating) tori, this leads to a stratification Cantor stratification.

**Chaos** An evolution of a dynamical system is chaotic if its future is badly predictable from its past. Examples of non-chaotic evolutions are periodic or multi-periodic. A system is called chaotic when many of its evolutions are. One criterion for chaoticity is the fact that one of the Lyapunov exponents is positive.

**Diophantine condition, Diophantine frequency vector** A frequency vector  $\omega \in \mathbb{R}^n$  is called Diophantine if there are constants  $\gamma > 0$  and  $\tau > n - 1$  with

$$|\langle k, \omega \rangle| \geq \frac{\gamma}{|k|^\tau} \quad \text{for all } k \in \mathbb{Z}^n \setminus \{0\}.$$

The Diophantine frequency vectors satisfying this condition for fixed  $\gamma$  and  $\tau$  form a Cantor set of half lines. As the Diophantine parameter  $\gamma$  tends to zero (while  $\tau$  remains fixed), these half lines extend to the origin. The complement in any compact set of frequency vectors satisfying a Diophantine condition with fixed  $\tau$  has a measure of order  $O(\gamma)$  as  $\gamma \downarrow 0$ .

**Integrable system** A Hamiltonian system with  $n$  degrees of freedom is (Liouville)-integrable if it has  $n$  functionally independent commuting. Locally this implies the existence of a torus

action, a feature that can be generalized to dissipative systems. In particular a mapping is integrable if it can be interpolated to become the stroboscopic mapping of a flow.

**KAM theory** Kolmogorov–Arnold–Moser theory is the perturbation theory of (Diophantine) quasi-periodic tori for nearly integrable Hamiltonian systems. In the format of quasi-periodic stability, the unperturbed and perturbed system, restricted to a Diophantine Cantor set, are smoothly conjugated in the sense of Whitney. This theory extends to the world of reversible, volume-preserving or general dissipative systems. In the latter KAM theory gives rise to families of quasi-periodic attractors. KAM theory also applies to torus bundles, in which case a global Whitney smooth conjugation can be proven to exist, that keeps track of the geometry. In an appropriate sense invariants like monodromy and Chern classes thus also can be defined in the nearly integrable case. Also compare with ► “Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite Dimensional Systems”.

**Nearly integrable system** In the setting of perturbation theory, a nearly integrable system is a perturbation of an integrable one. The latter then is an integrable approximation of the former. See an above item.

**Normal form truncation** Consider a dynamical system in the neighborhood of an equilibrium point, a fixed or periodic point, or a quasi-periodic torus, reducible to Floquet form. Then Taylor expansions (and their analogues) can be changed gradually into normal forms, that usually reflect the dynamics better. Often these display a (formal) torus symmetry, such that the normal form truncation becomes an integrable approximation, thus yielding a perturbation theory setting. See above items. Also compare with ► “Normal Forms in Perturbation Theory”.

**Persistent property** In the setting of perturbation theory, a property is persistent whenever it is inherited from the unperturbed to the perturbed system. Often the perturbation is taken in an appropriate topology on the space of systems, like the Whitney  $C^k$ -topology (Hirsch 1976).

**Perturbation problem** In perturbation theory the unperturbed systems usually are transparent

regarding their dynamics. Examples are integrable systems or normal form truncations. In a perturbation problem things are arranged in such a way that the original system is well-approximated by such an unperturbed one. This arrangement usually involves both changes of variables and scalings.

**Resonance** If the frequencies of an invariant torus with multi- or conditionally periodic flow are rationally dependent, this torus divides into invariant sub-tori. Such resonances  $\langle h, \omega \rangle = 0, h \in \mathbb{Z}^k$ , define hyperplanes in  $\omega$ -space and, by means of the frequency mapping, also in phase space. The smallest number  $|h| = |h_1| + \dots + |h_k|$  is the order of the resonance. Diophantine conditions describe a measure-theoretically large complement of a neighborhood of the (dense!) set of all resonances.

**Separatrices** Consider a hyperbolic equilibrium, fixed or periodic point or invariant torus. If the stable and unstable manifolds of such hyperbolic elements are codimension one immersed manifolds, then they are called separatrices, since they separate domains of phase space, for instance, basins of attraction.

**Singularity theory** A function  $H: \mathbb{R}^n \rightarrow \mathbb{R}$  has a critical point  $z \in \mathbb{R}^n$  where  $DH(z)$  vanishes. In local coordinates we may arrange  $z = 0$  (and similarly that it is mapped to zero as well). Two germs  $K: (\mathbb{R}^n, 0) \rightarrow (\mathbb{R}, 0)$  and  $N: (\mathbb{R}^n, 0) \rightarrow (\mathbb{R}, 0)$  represent the same function  $H$  locally around  $z$  if and only if there is a diffeomorphism  $\eta$  on  $\mathbb{R}^n$  satisfying

$$N = K \circ \eta.$$

The corresponding equivalence class is called a singularity.

**Structurally stable** A system is structurally stable if it is topologically equivalent to all nearby systems, where ‘nearby’ is measured in an appropriate topology on the space of systems, like the Whitney  $C^k$ -topology (Hirsch 1976). A family is structurally stable if for every nearby family there is a re-parametrization such that all corresponding systems are topologically equivalent.

## Definition of the Subject

The fundamental problem of mechanics is to study Hamiltonian systems that are small perturbations of integrable systems. Also, perturbations that destroy the Hamiltonian character are important, be it to study the effect of a small amount of friction, or to further the theory of dissipative systems themselves which surprisingly often revolves around certain well-chosen Hamiltonian systems. Furthermore there are approaches like KAM theory that historically were first applied to Hamiltonian systems. Typically perturbation theory explains only part of the dynamics, and in the resulting gaps the orderly unperturbed motion is replaced by random or chaotic motion.

## Introduction

We outline perturbation theory from a general point of view, illustrated by a few examples.

### The Perturbation Problem

The aim of perturbation theory is to approximate a given dynamical system by a more familiar one, regarding the former as a perturbation of the latter. The problem then is to deduce certain dynamical properties from the unperturbed to the perturbed case.

What is familiar may or may not be a matter of taste, at least it depends a lot on the dynamical properties of one's interest. Still the most frequently used unperturbed systems are:

- Linear systems
- Integrable Hamiltonian systems, compare with ► [“Dynamics of Hamiltonian Systems”](#) and references therein
- Normal form truncations, compare with ► [“Normal Forms in Perturbation Theory”](#) and references therein
- Etc.

To some extent the second category can be seen as a special case of the third. To avoid technicalities in this section we assume all systems to be sufficiently smooth, say of class  $C^\infty$  or real

analytic. Moreover in our considerations  $\varepsilon$  will be a real parameter. The unperturbed case always corresponds to  $\varepsilon = 0$  and the perturbed one to  $\varepsilon \neq 0$  or  $\varepsilon > 0$ .

**Examples of Perturbation Problems** To begin with consider the autonomous differential equation

$$\ddot{x} + \varepsilon \dot{x} + \frac{dV}{dx}(x) = 0,$$

modeling an oscillator with small damping. Rewriting this equation of motion as a planar vector field

$$\begin{aligned} \dot{x} &= -y \\ \dot{y} &= -\varepsilon y - \frac{dV}{dx}(x), \end{aligned}$$

we consider the energy  $H(x, y) = \frac{1}{2}y^2 + V(x)$ . For  $\varepsilon = 0$  the system is Hamiltonian with Hamiltonian function  $H$ . Indeed, generally we have  $\dot{H}(x, y) = -\varepsilon y^2$ , implying that for  $\varepsilon > 0$  there is dissipation of energy. Evidently for  $\varepsilon \neq 0$  the system is no longer Hamiltonian.

The reader is invited to compare the phase portraits of the cases  $\varepsilon = 0$  and  $\varepsilon > 0$  for  $V(x) = -\cos x$  (the pendulum) or  $V(x) = \frac{1}{2}\lambda x^2 + \frac{1}{24}bx^4$  (Duffing).

Another type of example is provided by the non-autonomous equation

$$\ddot{x} + \frac{dV}{dx}(x) = \varepsilon f(x, \dot{x}, t),$$

which can be regarded as the equation of motion of an oscillator with small external forcing. Again rewriting as a vector field, we obtain

$$\begin{aligned} \dot{t} &= 1 \\ \dot{x} &= -y \\ \dot{y} &= -\frac{dV}{dx}(x) + \varepsilon f(x, y, t), \end{aligned}$$

now on the generalized phase space  $\mathbb{R}^3 = \{t, x, y\}$ . In the case where the  $t$ -dependence is periodic, we can take  $\mathbb{S}^1 \times \mathbb{R}^2$  for (generalized) phase space.



**Remark**

- A small variation of the above driven system concerns a parametrically forced oscillator like

$$\ddot{x} + (\omega^2 + \varepsilon \cos t) \sin x = 0,$$

which happens to be entirely in the world of Hamiltonian systems.

- It may be useful to study the Poincaré or period mapping of such time periodic systems, which happens to be a mapping of the plane. We recall that in the Hamiltonian cases this mapping preserves area. For general reference in this direction see, e.g., Arnold (1978, 1983), Broer and Takens (2008), and Guckenheimer and Holmes (1983).

There are lots of variations and generalizations. One example is the solar system, where the unperturbed case consists of a number of uncoupled two-body problems concerning the Sun and each of the planets, and where the interaction between the planets is considered as small (Arnold 1978; Arnold and Avez 1967; Moser 1968, 1973).

**Remark**

- One variation is a restriction to fewer bodies, for example only three. Examples of this are systems like Sun–Jupiter–Saturn, Earth–Moon–Sun or Earth–Moon–Satellite.
- Often Sun, Moon and planets are considered as point masses, in which case the dynamics usually are modeled as a Hamiltonian system. It is also possible to extend this approach taking tidal effects into account, which have a non-conservative nature.
- The Solar System is close to resonance, which makes application of KAM theory problematic. There exist, however, other integrable approximations that take resonance into account (Arnold 1962; Féjoz 2004).

Quite another perturbation setting is local, e.g., near an equilibrium point. To fix thoughts consider

$$\dot{x} = Ax + f(x), \quad x \in \mathbb{R}^n$$

with  $A \in \mathfrak{gl}(n, \mathbb{R})$ ,  $f(0) = 0$ , and  $D_x f(0) = 0$ . By the scaling  $x = \varepsilon \bar{x}$  we rewrite the system to

$$\dot{\bar{x}} = A\bar{x} + \varepsilon g(\bar{x}).$$

So, here we take the linear part as an unperturbed system. Observe that for small  $\varepsilon$  the perturbation is small on a compact neighborhood of  $\bar{x} = 0$ .

This setting also has many variations. In fact, any normal form approximation may be treated in this way ▶ “Normal Forms in Perturbation Theory”. Then the normalized truncation forms the unperturbed part and the higher order terms the perturbation.

**Remark** In the above we took the classical viewpoint which involves a perturbation parameter controlling the size of the perturbation. Often one can generalize this by considering a suitable topology (like the Whitney topologies) on the corresponding class of systems (Hirsch 1976). Also compare with ▶ “Normal Forms in Perturbation Theory”, ▶ “Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite Dimensional Systems” and ▶ “Dynamics of Hamiltonian Systems”.

**Questions of Persistence**

What are the kind of questions perturbation theory asks? A large class of questions concerns the *persistence* of certain dynamical properties as known for the unperturbed case. To fix thoughts we give a few examples.

To begin with consider equilibria and periodic orbits. So we put

$$\dot{x} = f(x, \varepsilon), \quad x \in \mathbb{R}^n, \varepsilon \in \mathbb{R}, \quad (1)$$

for a map  $f: \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ . Recall that equilibria are given by the equation  $f(x, \varepsilon) = 0$ . The following theorem that continues equilibria in the unperturbed system for  $\varepsilon \neq 0$ , is a direct consequence of the implicit function theorem.

**Theorem 1 (Persistence of Equilibria)** Suppose that  $f(x_0, 0) = 0$  and that

$$D_x f(x_0, 0) \text{ has maximal rank.}$$

Then there exists a local arc  $\varepsilon \mapsto x(\varepsilon)$  with  $x(0) = x_0$  such that

$$f(x(\varepsilon), \varepsilon) \equiv 0.$$

Periodic orbits can be approximated in a similar way. Indeed, let the system (1) for  $\varepsilon = 0$  have a periodic orbit  $\gamma_0$ . Let  $\Sigma$  be a local transversal section of  $\gamma_0$  and  $P_0 : \Sigma \rightarrow \Sigma$  the corresponding Poincaré map. Then  $P_0$  has a fixed point  $x_0 \in \Sigma \cap \gamma_0$ . By transversality, for  $|\varepsilon|$  small, a local Poincaré map  $P_\varepsilon : \Sigma \rightarrow \Sigma$  is well-defined for Eq. (1). Observe that fixed points  $x_\varepsilon$  of  $P_\varepsilon$  correspond to periodic orbits  $\gamma_\varepsilon$  of Eq. (1). We now have, again as another direct consequence of the implicit function theorem.

**Theorem 2 (Persistence of Periodic Orbits)** In the above assume that

$$P_0(x_0) = x_0 \text{ and } D_x P_0(x_0) \text{ has no eigenvalue } 1.$$

Then there exists a local arc  $\varepsilon \mapsto x(\varepsilon)$  with  $x(0) = x_0$  such that

$$P_\varepsilon(x(\varepsilon)) \equiv x_\varepsilon.$$

#### Remark

- Often the conditions of Theorem 2 are not easy to verify. Sometimes it is useful here to use Floquet Theory, see Meyer and Hall (1992). In fact, if  $T_0$  is the period of  $\gamma_0$  and  $\Omega_0$  its Floquet matrix, then  $D_x P_0(x_0) = \exp(T_0 \Omega_0)$ .
- The format of the Theorems 1 and 2 with the perturbation parameter  $\varepsilon$  directly allows for algorithmic approaches. One way to proceed is by perturbation series, leading to asymptotic formulae that in the real analytic setting have positive radius of convergence. In the latter case the names of Poincaré and Lindstedt are associated with the method, cf. Arnol'd et al. (1988).

Also numerical continuation programmes exist based on the Newton method.

- The Theorems 1 and 2 can be seen as special cases of a general theorem for *normally hyperbolic invariant manifolds* (Hirsch et al. 1977, Theorem 4.1). In all cases a contraction

principle on a suitable Banach space of graphs leads to persistence of the invariant dynamical object.

This method in particular yields existence and persistence of stable and unstable manifolds (Chow and Hale 1982; Chow et al. 1994).

Another type of dynamics subject to perturbation theory is quasi-periodic. We emphasize that persistence of (Diophantine) quasi-periodic invariant tori occurs both in the conservative setting and in many others, like in the reversible and the general (dissipative) setting. In the latter case this leads to persistent occurrence of families of quasi-periodic attractors (Ruelle 1989). These results are in the domain of Kolmogorov–Arnold–Moser (KAM) theory. For details we refer to section “[KAM Theory: An Overview](#)” below or to Broer and Sevryuk (2007), ► “[Kolmogorov–Arnold–Moser \(KAM\) Theory for Finite and Infinite Dimensional Systems](#)”, the former reference containing more than 400 references in this area.

#### Remark

- Concerning the Solar System, KAM theory always has aimed at proving that it contains many quasi-periodic motions, in the sense of positive Liouville measure. This would imply that there is positive probability that a given initial condition lies on such a stable quasi-periodic motion (Arnold 1962; Féjóz 2004), however, also see Laskar (1995).
- Another type of result in this direction compares the distance of certain individual solutions of the perturbed and the unperturbed system, with coinciding initial conditions over time scales that are long in terms of  $\varepsilon$ . Compare with Broer and Sevryuk (2007).

Apart from persistence properties related to invariant manifolds or individual solutions, the aim can also be to obtain a more global persistence result. As an example of this we mention the Hartman–Grobman Theorem, e.g., Arnold (1983), Palis and de Melo (1982), and Robinson (1995). Here the setting once more is



$$\dot{x} = Ax + f(x), \quad x \in \mathbb{R}^n,$$

with  $A \in \mathfrak{gl}(n, \mathbb{R})$ ,  $f(0) = 0$  and  $D_x f(0) = 0$ . Now we assume  $A$  to be hyperbolic (i.e., with no purely imaginary eigenvalues). In that case the full system, near the origin, is topologically conjugated to the linear system  $\dot{x} = Ax$ . Therefore all global, *qualitative* properties of the unperturbed (linear) system are persistent under perturbation to the full system. For details on these notions see the above references, also compare with, e.g., Broer et al. (1991).

It is said that the hyperbolic linear system  $\dot{x} = Ax$  is (locally) *structurally stable*. This kind of thinking was introduced to the dynamical systems area by Thom (1989), with a first, successful application to catastrophe theory. For further details, see Arnold (1983), Broer et al. (1991), Hañßmann (2007), and Palis and de Melo (1982).

### General Dynamics

We give a few remarks on the general dynamics in a neighborhood of Hamiltonian KAM tori. In particular this concerns so-called superexponential stickiness of the KAM tori and adiabatic stability of the action variables, involving the so-called Nekhoroshev estimate.

To begin with, emphasize the following difference between the cases  $n = 2$  and  $n \geq 3$  in the classical KAM theorem of subsection “[Classical KAM Theory](#)”. For  $n = 2$  the level surfaces of the Hamiltonian are three-dimensional, while the Lagrangian tori have dimension two and hence codimension one in the energy hypersurfaces. This means that for open sets of initial conditions, the evolution curves are forever trapped in between KAM tori, as these tori foliate over nowhere dense sets of positive measure. This implies perpetual adiabatic stability of the action variables. In contrast, for  $n \geq 3$  the Lagrangian tori have codimension  $n - 1 > 1$  in the energy hypersurfaces and evolution curves may escape.

This actually occurs in the case of so-called *Arnold diffusion*. The literature on this subject is immense, and we here just quote (Arnold 1964; Arnold and Avez 1967; Marco and Sauzin 2003; Nekhoroshev 1977), for many more references see Broer and Sevryuk (2007).

Next we consider the motion in a neighborhood of the KAM tori, in the case where the systems are real analytic or at least Gevrey smooth. For a definition of Gevrey regularity see Wagener (2003). First we mention that, measured in terms of the distance to the KAM torus, nearby evolution curves generically stay nearby over a *superexponentially* long time (Morbidelli and Giorgilli 1995a, b). This property often is referred to as superexponential stickiness of the KAM tori, see Broer and Sevryuk (2007) for more references.

Second, nearly integrable Hamiltonian systems, in terms of the perturbation size, generically exhibit *exponentially* long adiabatic stability of the action variables, see e.g., Benettin (2005), Lochak (1999), Lochak and Marco (2005), Lochak and Neishtadt (1992), Marco and Sauzin (2003), Morbidelli and Giorgilli (1995b), Nekhoroshev (1977, 1985), Niederman (2004), and Pöschel (1993), ► “[Nekhoroshev Theory](#)” and many others, for more references see Broer and Sevryuk (2007). This property is referred to as the *Nekhoroshev estimate* or the *Nekhoroshev theorem*. For related work on perturbations of so-called superintegrable systems, also see Broer and Sevryuk (2007) and references therein.

### Chaos

In the previous subsection we discussed persistent and some non-persistent features of dynamical systems under small perturbations. Here we discuss properties related to splitting of separatrices, caused by generic perturbations.

A first example was met earlier, when comparing the pendulum with and without (small) damping. The unperturbed system is the undamped one and this is a Hamiltonian system. The perturbation however no longer is Hamiltonian. We see that the equilibria are persistent, as they should be according to Theorem 1, but that none of the periodic orbits survives the perturbation. Such qualitative changes go with perturbing away from the Hamiltonian setting.

Similar examples concern the breaking of a certain symmetry by the perturbation. The latter often occurs in the case of normal form approximations. Then the normalized truncation is viewed as the unperturbed system, which is

perturbed by the higher order terms. The truncation often displays a reasonable amount of symmetry (e.g., toroidal symmetry), which *generically* is forbidden for the class of systems under consideration, e.g., see Broer and Takens (1989).

To fix thoughts we reconsider the conservative example

$$\ddot{x} + (\omega^2 + \varepsilon \cos t) \sin x = 0$$

of the previous section. The corresponding (time dependent, Hamiltonian (Arnold 1978)) vector field reads

$$\begin{aligned} \dot{t} &= 1 \\ \dot{x} &= y \\ \dot{y} &= -(\omega^2 + \varepsilon \cos t) \sin x. \end{aligned}$$

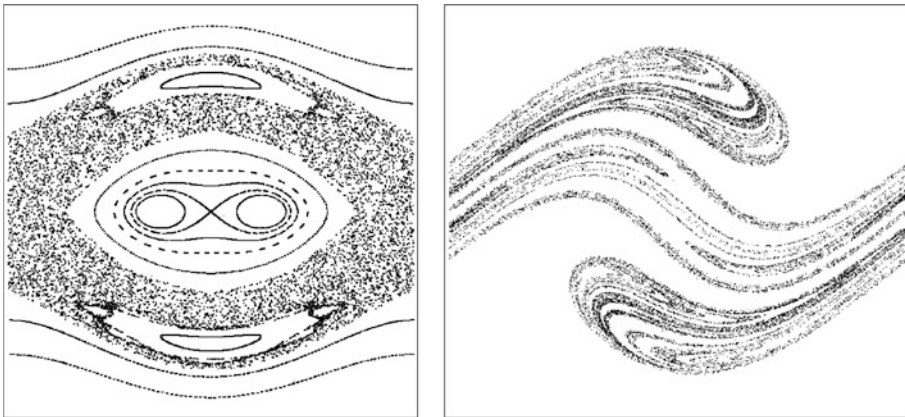
Let  $P_{\omega, \varepsilon}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the corresponding (area-preserving) Poincaré map. Let us consider the unperturbed map  $P_{\omega, 0}$  which is just the flow over time  $2\pi$  of the free pendulum  $\ddot{x} + \omega^2 \sin x = 0$ . Such a map is called *integrable*, since it is the stroboscopic map of a two-dimensional vector field, hence displaying the  $\mathbb{R}$ -symmetry of a flow. When perturbed to the *nearly integrable* case  $\varepsilon \neq 0$ , this symmetry generically is broken. We list a few of the generic properties for such maps (Robinson 1995):

- The homoclinic and heteroclinic points occur at transversal intersections of the corresponding stable and unstable manifolds.
- The periodic points of period less than a given bound are isolated.

This means generically that the separatrices split and that the resonant invariant circles filled with periodic points with the same (rational) rotation number fall apart. In any concrete example the issue remains whether or not it satisfies appropriate genericity conditions. One method to check this is due to Melnikov, compare Guckenheimer and Holmes (1983) and Wiggins (1990), for more sophisticated tools see Gelfreich and Lazutkin (2001). Often this leads to elliptic (Abelian) integrals.

In nearly integrable systems chaos can occur. This fact is at the heart of the celebrated non-integrability of the three-body problem as addressed by Poincaré (Barrow-Green 1997; Diacu and Holmes 1996; Moser 1968, 1973; Poincaré 1980). A long standing open conjecture is that the clouds of points as visible in Fig. 1, left, densely fill sets of positive area, thereby leading to ergodicity (Arnold and Avez 1967).

In the case of dissipation, see Fig. 1, right, we conjecture the occurrence of a Hénon-like strange attractor (Benedicks and Carleson 1991; Broer and Krauskopf 2000; Ruelle and Takens 1971).



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 1** Chaos in the parametrically forced pendulum. (Left) Poincaré map  $P_{\omega, \varepsilon}$  near the  $1:2$  resonance  $\omega = \frac{1}{2}$  and for  $\varepsilon > 0$  not too small. (Right) A dissipative analogue

**Remark**

- The persistent occurrence of periodic points of a given rotation number follows from the Poincaré–Birkhoff fixed point theorem (Hofer and Zehnder 1994; McDuff and Salamon 1995; Moser 1968), i.e., on topological grounds.
- The above arguments are not restricted to the conservative setting, although quite a number of unperturbed systems come from this world. Again see Fig. 1.

**One Degree of Freedom**

Planar Hamiltonian systems are always integrable and the orbits are given by the level sets of the Hamiltonian function. This still leaves room for a perturbation theory. The recurrent dynamics consists of periodic orbits, equilibria and asymptotic trajectories forming the (un)stable manifolds of unstable equilibria. The equilibria organize the phase portrait, and generically all equilibria are elliptic (purely imaginary eigenvalues) or hyperbolic (real eigenvalues), i.e. there is no equilibrium with a vanishing eigenvalue. If the system depends on a parameter such vanishing eigenvalues may be unavoidable and it becomes possible that the corresponding dynamics persist under perturbations.

Perturbations may also destroy the Hamiltonian character of the flow. This happens especially where the starting point is a dissipative planar system and e.g. a scaling leads for  $\varepsilon = 0$  to a limiting Hamiltonian flow. The perturbation problem then becomes twofold. Equilibria still persist by Theorem 1 and hyperbolic equilibria moreover persist as such, with the sum of eigenvalues of order  $O(\varepsilon)$ . Also for elliptic eigenvalues the sum of eigenvalues is of order  $O(\varepsilon)$  after the perturbation, but here this number measures the dissipation whence the equilibrium becomes (weakly) attractive for negative values and (weakly) unstable for positive values. The one-parameter families of periodic orbits of a Hamiltonian system do not persist under dissipative perturbations, the very fact that they form families imposes the corresponding fixed point of the Poincaré mapping to have an eigenvalue one and Theorem 2 does not apply. Typically only finitely many periodic orbits survive a dissipative

perturbation and it is already a difficult task to determine their number.

**Hamiltonian Perturbations**

The Duffing oscillator has the Hamiltonian function

$$H(x, y) = \frac{1}{2}y^2 + \frac{1}{24}bx^4 + \frac{1}{2}\lambda x^2 \quad (2)$$

where  $b$  is a constant distinguishing the two cases  $b = \pm 1$  and  $\lambda$  is a parameter. Under variation of the parameter the equations of motion

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= -\frac{1}{6}bx^3 - \lambda x \end{aligned}$$

display a Hamiltonian pitchfork bifurcation, supercritical for positive  $b$  and subcritical in case  $b$  is negative. Correspondingly, the linearization at the equilibrium  $x = 0$  of the anharmonic oscillator  $\lambda = 0$  is given by the matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

whence this equilibrium is parabolic.

The typical way in which a parabolic equilibrium bifurcates is the center-saddle bifurcation. Here the Hamiltonian reads

$$H(x, y) = \frac{1}{2}ay^2 + \frac{1}{6}bx^3 + c\lambda x \quad (3)$$

where  $a, b, c \in \mathbb{R}$  are nonzero constants, for instance  $a = b = c = 1$ . Note that this is a completely different unfolding of the parabolic equilibrium at the origin. A closer look at the phase portraits and in particular at the Hamiltonian function of the Hamiltonian pitchfork bifurcation reveals the symmetry  $x \mapsto -x$  of the Duffing oscillator. This suggests the addition of the non-symmetric term  $\mu x$ . The resulting two-parameter family

$$H_{\lambda, \mu}(x, y) = \frac{1}{2}y^2 + \frac{1}{24}bx^4 + \frac{1}{2}\lambda x^2 + \mu x$$

of Hamiltonian systems is indeed structurally stable. This implies not only that all equilibria of a

Hamiltonian perturbation of the Duffing oscillator have a local flow equivalent to the local flow near a suitable equilibrium in this two-parameter family, but that every one-parameter family of  $\mathbb{Z}_2$ -symmetric Hamiltonian systems that is a perturbation of Eq. (2) has equivalent dynamics. For more details see Broer et al. (2003c) and references therein.

This approach applies *mutatis mutandis* to every non-degenerate planar singularity, cf. Hanßmann (2007) and Takens (1973). At an equilibrium all partial derivatives of the Hamiltonian vanish and the resulting singularity is called non-degenerate if it has finite multiplicity, which implies that it admits a versal unfolding  $H_\lambda$  with finitely many parameters. The family of Hamiltonian systems defined by this versal unfolding contains all possible (local) dynamics that the initial equilibrium may be perturbed to. Imposing additional discrete symmetries is immediate, the necessary symmetric versal unfolding is obtained by averaging

$$H_\lambda^G = \frac{1}{|G|} \sum_{g \in G} H_\lambda \circ g$$

along the orbits of the symmetry group  $G$ .

### Dissipative Perturbations

In a generic dissipative system all equilibria are hyperbolic. Qualitatively, i.e. up to topological equivalence, the local dynamics is completely determined by the number of eigenvalues with positive real part. Those hyperbolic equilibria that can appear in Hamiltonian systems (the eigenvalues forming pairs  $\pm i\nu$ ) do not play an important role. Rather, planar Hamiltonian systems become important as a tool to understand certain bifurcations triggered off by non-hyperbolic equilibria. Again this requires the system to depend on external parameters.

The simplest example is the Hopf bifurcation, a co-dimension one bifurcation where an equilibrium loses stability as the pair of eigenvalues crosses the imaginary axis, say at  $\pm i$ . At the bifurcation the linearization is a Hamiltonian system with an elliptic equilibrium (the co-dimension one bifurcations where a single eigenvalue crosses the

imaginary axis through 0 do not have a Hamiltonian linearization). This limiting Hamiltonian system has a one-parameter family of periodic orbits around the equilibrium, and the non-linear terms determine the fate of these periodic orbits. The normal form of order three reads

$$\begin{aligned}\dot{x} &= -y(1 + b(x^2 + y^2)) + x(\lambda + a(x^2 + y^2)) \\ \dot{y} &= x(1 + b(x^2 + y^2)) + y(\lambda + a(x^2 + y^2))\end{aligned}$$

and is Hamiltonian if and only if  $(\lambda, a) = (0, 0)$ . The sign of the coefficient distinguishes between the supercritical case  $a > 0$ , in which there are no periodic orbits coexisting with the attractive equilibria (i.e. when  $\lambda < 0$ ) and one attracting periodic orbit for each  $\lambda > 0$  (coexisting with the unstable equilibrium), and the subcritical case  $a < 0$ , in which the family of periodic orbits is unstable and coexists with the attractive equilibria (with no periodic orbits for parameters  $\lambda > 0$ ). As  $\lambda \rightarrow 0$  the family of periodic orbits shrinks down to the origin, so also this Hamiltonian feature is preserved.

Equilibria with a double eigenvalue 0 need two parameters to persistently occur in families of dissipative systems. The generic case is the Takens–Bogdanov bifurcation. Here the linear part is too degenerate to be helpful, but the nonlinear Hamiltonian system defined by Eq. (3) with  $a = 1 = c\lambda$  and  $b = -3$  provides the periodic and heteroclinic orbit(s) that constitute the nontrivial part of the bifurcation diagram. Where discrete symmetries are present, e.g. for equilibria in dissipative systems originating from other generic bifurcations, the limiting Hamiltonian system exhibits that same discrete symmetry. For more details see Chow et al. (1994), Guckenheimer and Holmes (1983), Kuznetsov (2004), and references therein.

The continuation of certain periodic orbits from an unperturbed Hamiltonian system under dissipative perturbation can be based on Melnikov-like methods, again see Guckenheimer and Holmes (1983) and Wiggins (1990). As above, this often leads to Abelian integrals, for instance to count the number of periodic orbits that branch off.

### Reversible Perturbations

A dynamical system that admits a reflection symmetry  $R$  mapping trajectories  $\varphi(t, z_0)$  to

trajectories  $\varphi(-t, R(z_0))$  is called reversible. In the planar case we may restrict to the reversing reflection

$$\begin{aligned} R : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (x, -y). \end{aligned} \quad (4)$$

All Hamiltonian functions  $H = \frac{1}{2}y^2 + V(x)$  which have an interpretation “kinetic + potential energy” are reversible, and in general the class of reversible systems is positioned between the class of Hamiltonian systems and the class of dissipative systems. A guiding example is the perturbed Duffing oscillator (with the roles of  $x$  and  $y$  exchanged so that Eq. (4) remains the reversing symmetry)

$$\begin{aligned} \dot{x} &= -\frac{1}{6}y^3 - y + \varepsilon xy \\ \dot{y} &= x \end{aligned}$$

that combines the Hamiltonian character of the equilibrium at the origin with the dissipative character of the two other equilibria. Note that all orbits outside the homoclinic loop are periodic.

There are two ways in which the reversing symmetry (4) imposes a Hamiltonian character on the dynamics. An equilibrium that lies on the symmetry line  $\{y = 0\}$  has a linearization that is itself a reversible system and consequently the eigenvalues are subject to the same constraints as in the Hamiltonian case. (For equilibria  $z_0$  that do not lie on the symmetry line the reflection  $R(z_0)$  is also an equilibrium, and it is to the union of their eigenvalues that these constraints still apply.) Furthermore, every orbit that crosses  $\{y = 0\}$  more than once is automatically periodic, and these periodic orbits form one-parameter families. In particular, elliptic equilibria are still surrounded by periodic orbits.

The dissipative character of a reversible system is most obvious for orbits that do not cross the symmetry line. Here  $R$  merely maps the orbit to a reflected counterpart. The above perturbed Duffing oscillator exemplifies that the character of an orbit crossing  $\{y = 0\}$  exactly once is undetermined. While the homoclinic orbit of the saddle at the origin has a Hamiltonian character,

the heteroclinic orbits between the other two equilibria behave like in a dissipative system.

## Perturbations of Periodic Orbits

The perturbation of a one-degree-of-freedom system by a periodic forcing is a perturbation that changes the phasespace. Treating the time variable  $t$  as a phase space variable leads to the extended phase space  $\mathbb{S}^1 \times \mathbb{R}^2$  and equilibria of the unperturbed system become periodic orbits, inheriting the normal behavior. Furthermore introducing an action conjugate to the “angle”  $t$  yields a Hamiltonian system in two degrees of freedom.

While the one-parameter families of periodic orbits merely provide the typical recurrent motion in one degree of freedom, they form special solutions in two or more degrees of freedom. Arcs of elliptic periodic orbits are particularly instructive. Note that these occur generically in both the Hamiltonian and the reversible context.

### Conservative Perturbations

Along the family of elliptic periodic orbits a pair  $e^{\pm i\Omega}$  of Floquet multipliers passes regularly through roots of unity. Generically this happens on a dense set of parameter values, but for fixed denominator  $q$  in  $e^{\pm i\Omega} = e^{\pm 2\pi i p/q}$  the corresponding energy values are isolated. The most important of such resonances are those with small denominators  $q$ .

For  $q = 1$  generically a periodic center-saddle bifurcation takes place where an elliptic and a hyperbolic periodic orbit meet at a parabolic periodic orbit. No periodic orbit remains under further variation of a suitable parameter.

The generic bifurcation for  $q = 2$  is the period-doubling bifurcation where an elliptic periodic orbit turns hyperbolic (or vice versa) when passing through a parabolic periodic orbit with Floquet multipliers  $-1$ . Furthermore, a family of periodic orbits with twice the period emerges from the parabolic periodic orbit, inheriting the normal linear behavior from the initial periodic orbit.

In case  $q = 3$ , and possibly also for  $q = 4$ , generically two arcs of hyperbolic periodic orbits emerge, both with three (resp. four) times the



period. One of these extends for lower and the other for higher parameter values. The initial elliptic periodic orbit momentarily loses its stability due to these approaching unstable orbits.

Denominators  $q \geq 5$  (and also the second possibility for  $q = 4$ ) lead to a pair of subharmonic periodic orbits of  $q$  times the period emerging either for lower or for higher parameter values. This is (especially for large  $q$ ) comparable to the behavior at Diophantine  $e^{\pm i\Omega}$  where a family of invariant tori emerges, cf. section “[Invariant Curves of Planar Diffeomorphisms](#)” below.

For a single pair  $e^{\pm i\Omega}$  of Floquet multipliers this behavior is traditionally studied for the iso-energetic Poincaré-mapping, cf. MacKay (1993) and references therein. However, the above description remains true in higher dimensions, where additionally multiple pairs of Floquet multipliers may interact. An instructive example is the Lagrange top, the sleeping motion of which is gyroscopically stabilized after a periodic Hamiltonian Hopf bifurcation; see Cushman and Bates (1997) for more details.

### Dissipative Perturbations

There exists a large class of local bifurcations in the dissipative setting, that can be arranged in a perturbation theory setting, where the unperturbed system is Hamiltonian. The arrangement consists of changes of variables and rescaling. An early example of this is the Bogdanov–Takens bifurcation (Takens 1974a, b). For other examples regarding nilpotent singularities, see Broer and Roussarie (2001), Broer et al. (2006c), and references therein.

To fix thoughts, consider families of planar maps and let the unperturbed Hamiltonian part contain a center (possibly surrounded by a homoclinic loop). The question then is which of these persist when adding the dissipative perturbation.

Usually only a definite finite number persists. As in subsection “[Chaos](#)”, a Melnikov function can be invoked here, possibly again leading to elliptic (Abelian) integrals, Picard Fuchs equations, etc. For details see Dumortier et al. (1991), Roussarie (1997), and references therein.

## Invariant Curves of Planar Diffeomorphisms

This section starts with general considerations on circle diffeomorphisms, in particular focusing on persistence properties of quasi-periodic dynamics. Our main references are Arnold (1961), Broer and Sevryuk (2007), Broer et al. (1990, 1996), Herman (1977, 1979), and Yoccoz (1983, 1992). For a definition of rotation number, see Devaney (1989). After this we turn to area preserving maps of an annulus where we discuss Moser’s twist map theorem (Moser 1962; also see Broer and Sevryuk 2007; Broer et al. (1990, 1996)). The section is concluded by a description of the holomorphic linearization of a fixed point in a planar map (Arnold 1983; Milnor 2006; Yoccoz 1995, 2002).

Our main perspective will be perturbative, where we consider circle maps near a rigid rotation. It turns out that generally parameters are needed for persistence of quasi-periodicity under perturbations. In the area preserving setting we consider perturbations of a pure twist map.

### Circle Maps

We start with the following general problem. Given a two-parameter family

$$P_{\alpha,\varepsilon} : \mathbb{T}^1 \rightarrow \mathbb{T}^1, \quad x \mapsto x + 2\pi\alpha + \varepsilon a(x, \alpha, \varepsilon)$$

of circle maps of class  $C^\infty$ . It turns out to be convenient to view this two-parameter family as a one-parameter family of maps

$$\begin{aligned} P_\varepsilon : \mathbb{T}^1 \times [0, 1] &\rightarrow \mathbb{T}^1 \times [0, 1], \\ (x, \alpha) &\mapsto (x + 2\pi\alpha + \varepsilon a(x, \alpha, \varepsilon), \alpha) \end{aligned}$$

of the cylinder. Note that the unperturbed system  $p_0$  is a family of rigid circle rotations, viewed as a cylinder map, where the individual map  $P_{\alpha,0}$  has rotation number  $\alpha$ . The question now is what will be the fate of this rigid dynamics for  $0 \neq |\varepsilon| \ll 1$ .

The classical way to address this question is to look for a conjugation  $\Phi_\varepsilon$ , that makes the following diagram commute

$$\begin{array}{ccc} \mathbb{T}^1 \times [0, 1] & \xrightarrow{P_\varepsilon} & \mathbb{T}^1 \times [0, 1] \\ \uparrow \Phi_\varepsilon & & \uparrow \Phi_\varepsilon \\ \mathbb{T}^1 \times [0, 1] & \xrightarrow{P_0} & \mathbb{T}^1 \times [0, 1], \end{array}$$

i.e., such that

$$P_\varepsilon \circ \Phi_\varepsilon = \Phi_\varepsilon \circ P_0.$$

Due to the format of  $P_\varepsilon$  we take  $\Phi_\varepsilon$  as a skew map

$$\Phi_\varepsilon(x, \alpha) = (x + \varepsilon U(x, \alpha, \varepsilon), \alpha + \varepsilon \sigma(\alpha, \varepsilon)),$$

which leads to the *nonlinear* equation

$$\begin{aligned} U(x + 2\pi\alpha, \alpha, \varepsilon) - U(x, \alpha, \varepsilon) \\ = 2\pi\sigma(\alpha, \varepsilon) + a(x + \varepsilon U(x, \alpha, \varepsilon), \alpha + \varepsilon\sigma(\alpha, \varepsilon), \varepsilon) \end{aligned}$$

in the unknown maps  $U$  and  $\sigma$ . Expanding in powers of  $\varepsilon$  and comparing at lowest order yields the linear equation

$$\begin{aligned} U_0(x + 2\pi\alpha, \alpha) - U_0(x, \alpha) \\ = 2\pi\sigma_0(\alpha) + a_0(x, \alpha) \end{aligned}$$

which can be directly solved by Fourier series. Indeed, writing

$$\begin{aligned} a_0(x, \alpha) &= \sum_{k \in \mathbb{Z}} a_{0k}(\alpha) e^{ikx}, \\ U_0(x, \alpha) &= \sum_{k \in \mathbb{Z}} U_{0k}(\alpha) e^{ikx} \end{aligned}$$

we find  $\sigma_0 = -1/(2\pi)a_{00}$  and

$$U_{0k}(\alpha) = \frac{a_{0k}(\alpha)}{e^{2\pi i k \alpha} - 1}.$$

It follows that in general a formal solution exists if and only if  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ . Still, the accumulation of  $e^{2\pi i k \alpha} - 1$  on 0 leads to the celebrated *small divisors* (Arnold and Avez 1967; Moser 1973; also see Broer and Sevryuk 2007; Broer et al. 1990, 1996; Ciocci et al. 2005).

The classical solution considers the following Diophantine non-resonance conditions. Fixing

$\tau > 2$  and  $\gamma > 0$  consider  $\alpha \in [0, 1]$  such that for all rationals  $p/q$

$$\left| \alpha - \frac{p}{q} \right| \geq \gamma q^{-\tau}. \quad (5)$$

This subset of such  $\alpha$ s is denoted by  $[0, 1]_{\tau, \gamma}$  and is well-known to be nowhere dense but of large measure as  $\gamma > 0$  gets small (Oxtoby 1971). Note that Diophantine numbers are irrational.

**Theorem 3 (Circle Map Theorem)** For  $\gamma$  sufficiently small and for the perturbation  $\varepsilon a$  sufficiently small in the  $C^\infty$ -topology, there exists a  $C^\infty$  transformation  $\Phi_\varepsilon : \mathbb{T}^1 \times [0, 1] \rightarrow \mathbb{T}^1 \times [0, 1]$ , conjugating the restriction  $P_0|_{[0, 1]_{\tau, \gamma}}$  to a subsystem of  $P_\varepsilon$ .

Theorem 3 in the present structural stability formulation (compare with Fig. 2) is a special case of the results in Broer et al. (1990, 1996). We here speak of *quasi-periodic stability*. For earlier versions see Arnold (1961) and Arnold and Avez (1967).

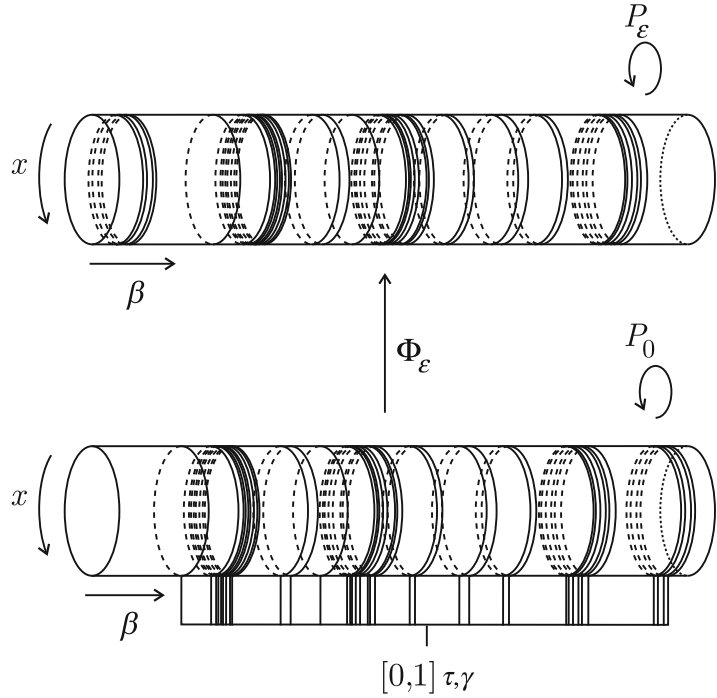
#### Remark

- Rotation numbers are preserved by the map  $\Phi_\varepsilon$  and irrational rotation numbers correspond to quasi-periodicity. Theorem 3 thus ensures that *typically* quasi-periodicity occurs with *positive* measure in the parameter space. Note that since Cantor sets are perfect, quasi-periodicity typically has a non-isolated occurrence.
- The map  $\Phi_\varepsilon$  has no dynamical meaning inside the gaps. The gap dynamics in the case of circle maps can be illustrated by the Arnold family of circle maps Arnold (1961, 1983; Devaney 1989), given by

$$P_{\alpha, \varepsilon}(x) = x + 2\pi\alpha + \varepsilon \sin x$$

- which exhibits a countable union of open resonance tongues where the dynamics is periodic, see Fig. 3. Note that this map only is a diffeomorphism for  $|\varepsilon| < 1$ .
- We like to mention that non-perturbative versions of Theorem 3 have been proven in Herman (1977, 1979) and Yoccoz (1983).

**Hamiltonian  
Perturbation Theory  
(and Transition to  
Chaos), Fig. 2** Skew  
cylinder map, conjugating  
(Diophantine) quasi-  
periodic invariant circles of  
 $P_0$  and  $P_\varepsilon$



- For simplicity we formulated Theorem 3 under  $C^\infty$ -regularity, noting that there exist many ways to generalize this. On the one hand there exist  $C^k$ -versions for finite  $k$  and on the other hand there exist fine tunings in terms of real-analytic and Gevrey regularity. For details we refer to Broer and Sevryuk (2007), Broer et al. (1996), and references therein. This same remark applies to other results in this section and in section “KAM Theory: An Overview” on KAM theory.

A possible application of Theorem 3 runs as follows. Consider a system of weakly coupled Van der Pol oscillators

$$\begin{aligned}\ddot{y}_1 + c_1 \dot{y}_1 + a_1 y_1 + f_1(y_1, \dot{y}_1) &= \varepsilon g_1(y_1, y_2, \dot{y}_1, \dot{y}_2) \\ \ddot{y}_2 + c_2 \dot{y}_2 + a_2 y_2 + f_2(y_2, \dot{y}_2) &= \varepsilon g_2(y_1, y_2, \dot{y}_1, \dot{y}_2).\end{aligned}$$

Writing  $\dot{y}_j = z_j, j = 1, 2$ , one obtains a vector field in the four-dimensional phase space  $\mathbb{R}^2 \times \mathbb{R}^2 = \{(y_1, z_1), (y_2, z_2)\}$ . For  $\varepsilon = 0$  this vector field has an invariant two-torus, which is the product of the periodic motions of the individual Van der Pol oscillations. This two-torus is normally

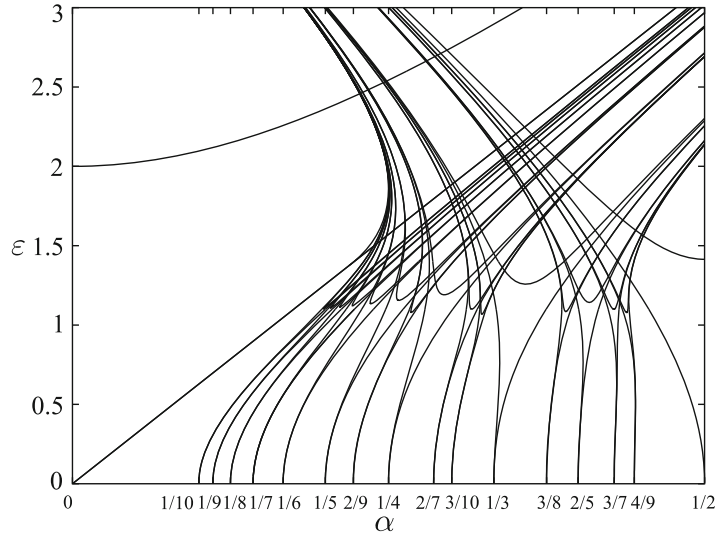
hyperbolic and therefore persistent for  $|\varepsilon| \ll 1$  (Hirsch et al. 1977). In fact the torus is an attractor and we can define a Poincaré return map within this torus attractor. If we include some of the coefficients of the equations as parameters, Theorem 3 is directly applicable. The above statements on quasi-periodic circle maps then directly translate to the case of quasi-periodic invariant two-tori. Concerning the resonant cases, generically a tongue structure like in Fig. 3 occurs; for the dynamics corresponding to parameter values inside such a tongue one speaks of *phase lock*.

#### Remark

- The celebrated synchronization of Huygens’ clocks (1888) is related to a 1 : 1 resonance, meaning that the corresponding Poincaré map would have its parameters in the main tongue with rotation number 0. Compare with Fig. 3.
- There exist direct generalizations to cases with  $n$ -oscillators ( $n \in \mathbb{N}$ ), leading to families of invariant  $n$ -tori carrying quasi-periodic flow, forming a nowhere dense set of positive measure. An alteration with resonance occurs as roughly sketched in Fig. 3. In higher



**Hamiltonian  
Perturbation Theory  
(and Transition to  
Chaos), Fig. 3** Arnold  
resonance tongues; for  $\varepsilon \geq 1$   
the maps are endomorphic



dimension the gap dynamics, apart from periodicity, also can contain strange attractors (Newhouse et al. 1978; Ruelle and Takens 1971). We shall come back to this subject in a later section.

### Area-Preserving Maps

The above setting historically was preceded by an area preserving analogue (Moser 1962) that has its origin in the Hamiltonian dynamics of frictionless mechanics.

Let  $\Delta \subseteq \mathbb{R}^2 \setminus \{(0, 0)\}$  be an annulus, with symplectic polar coordinates  $(\varphi, I) \in \mathbb{T}^1 \times \mathbf{K}$ , where  $\mathbf{K}$  is an interval. Moreover, let  $\sigma = d\varphi \wedge dI$  be the area form on  $\Delta$ .

We consider a  $\sigma$ -preserving smooth map  $P_\varepsilon : \Delta \rightarrow \Delta$  of the form

$$P_\varepsilon(\varphi, I) = (\varphi + 2\pi\alpha(I), I) + O(\varepsilon),$$

where we assume that the map  $I \mapsto \alpha(I)$  is a (local) diffeomorphism. This assumption is known as the *twist condition* and  $P_\varepsilon$  is called a *twist map*. For the unperturbed case  $\varepsilon = 0$  we are dealing with a pure twist map and its dynamics are comparable to the unperturbed family of cylinder maps as met in subsection “Circle Maps”. Indeed it is again a family of rigid rotations, parametrized by  $I$  and where  $P_0(\cdot, I)$  has rotation number  $\alpha(I)$ . In this case the question is what will be the fate of

this family of invariant circles, as well as with the corresponding rigidly rotational dynamics.

Regarding the rotation number we again introduce Diophantine conditions. Indeed, for  $\tau > 2$  and  $\gamma > 0$  the subset  $[0, 1]_{\tau, \gamma}$  is defined as in Eq. (5), i.e., it contains all  $\alpha \in [0, 1]$ , such that for all rationals  $p/q$

$$\left| \alpha - \frac{p}{q} \right| \geq \gamma q^{-\tau}.$$

Pulling back  $[0, 1]_{\tau, \gamma}$  along the map  $\alpha$  we obtain a subset  $\Delta_{\tau, \gamma} \subseteq \Delta$ .

**Theorem 4 (Twist Map Theorem (Moser 1962))** For  $\gamma$  sufficiently small, and for the perturbation  $O(\varepsilon)$  sufficiently small in  $C^\infty$ -topology, there exists a  $C^\infty$  transformation  $\Phi_\varepsilon : \Delta \rightarrow \Delta$ , conjugating the restriction  $P_0|_{\Delta_{\tau, \gamma}}$  to a subsystem of  $P_\varepsilon$ .

As in the case of Theorem 3 again we chose the formulation of Broer et al. (1990, 1996). Largely the remarks following Theorem 3 also apply here.

### Remark

- Compare the format of the Theorems 3 and 4 and observe that in the latter case the role of the parameter  $\alpha$  has been taken by the action variable  $I$ . Theorem 4 implies that *typically* quasi-periodicity occurs with positive measure in phase space.

- In the gaps typically we have coexistence of periodicity, quasi-periodicity and chaos (Arnold 1978; Arnold and Avez 1967; Broer et al. 2003b; Moser 1968, 1973; Robinson 1995; Wiggins 1990). The latter follows from transversality of homo- and heteroclinic connections that give rise to positive topological entropy. Open problems are whether the corresponding Lyapunov exponents also are positive, compare with the discussion at the end of the introduction.

Similar to the applications of Theorem 3 given at the end of subsection “Circle Maps”, here direct applications are possible in the conservative setting. Indeed, consider a system of weakly coupled pendula

$$\begin{aligned}\ddot{y}_1 + \alpha_1^2 \sin y_1 &= \varepsilon \frac{\partial U}{\partial y_1}(y_1, y_2) \\ \ddot{y}_2 + \alpha_2^2 \sin y_2 &= \varepsilon \frac{\partial U}{\partial y_2}(y_1, y_2).\end{aligned}$$

Writing  $\dot{y}_j = z_j, j = 1, 2$  as before, we again get a vector field in the four-dimensional phase space  $\mathbb{R}^2 \times \mathbb{R}^2 = \{(y_1, y_2), (z_1, z_2)\}$ . In this case the energy

$$\begin{aligned}H_\varepsilon(y_1, y_2, z_1, z_2) \\ = \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2 - \alpha_1^2 \cos y_1 - \alpha_2^2 \cos y_2 + \varepsilon U(y_1, y_2)\end{aligned}$$

is a constant of motion. Restricting to a three-dimensional energy surface  $H_\varepsilon^{-1} = \text{const.}$ , the iso-energetic Poincaré map  $P_\varepsilon$  is a twist map and application of Theorem 4 yields the conclusion of quasi-periodicity (on invariant two-tori) occurring with positive measure in the energy surfaces of  $H_\varepsilon$ .

**Remark** As in the dissipative case this example directly generalizes to cases with  $n$  oscillators ( $n \in \mathbb{N}$ ), again leading to invariant  $n$ -tori with quasi-periodic flow. We shall return to this subject in a later section.

### Linearization of Complex Maps

The subsections “Circle Maps” and “Area-Preserving Maps” both deal with smooth circle maps that are conjugated to rigid rotations.

Presently the concern is with planar holomorphic maps that are conjugated to a rigid rotation on an open subset of the plane. Historically this is the first time that a small divisor problem was solved (Arnold 1983; Milnor 2006; Yoccoz 1995, 2002) and ► “Convergence of Perturbative Expansions”.

### Complex Linearization

Given is a holomorphic germ  $F : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$  of the form  $F(z) = \lambda z + f(z)$ , with  $f(0) = f'(0) = 0$ . The problem is to find a biholomorphic germ  $\Phi : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$  such that

$$\Phi \circ F = \lambda \cdot \Phi.$$

Such a diffeomorphism  $\Phi$  is called a *linearization* of  $F$  near 0.

We begin with the formal approach. Given the series  $f(z) = \sum_{j \geq 2} f_j z^j$ , we look for  $\Phi(z) = z + \sum_{j \geq 2} \phi_j z^j$ . It turns out that a solution always exists whenever  $\lambda \neq 0$  is not a root of unity. Indeed, direct computation reveals the following set of equations that can be solved recursively:

For **j = 2**: get the equation  $\lambda(1 - \lambda)\phi_2 = f_2$

For **j = 3**: get the equation  $\lambda(1 - \lambda^2)\phi_3 = f_3 + 2\lambda f_2 \phi_2$

For **j = n**: get the equation  $\lambda(1 - \lambda^{n-1})\phi_n = f_n + \text{known.}$

The question now reduces to whether this formal solution has a positive radius of convergence.

The hyperbolic case  $0 < |\lambda| \neq 1$  was already solved by Poincaré, for a description see Arnold (1983). The elliptic case  $|\lambda| = 1$  again has small divisors and was solved by Siegel when for some  $\gamma > 0$  and  $\tau > 2$  we have the Diophantine non-resonance condition

$$|\lambda - e^{2\pi i q}| \geq \gamma |q|^{-\tau}.$$

The corresponding set of  $\lambda$  constitutes a set of full measure in  $\mathbb{T}^1 = \{\lambda\}$ .

Yoccoz (1995) completely solved the elliptic case using the Bruno-condition. If

$$\lambda = e^{2\pi i \alpha} \text{ and when } \frac{p_n}{q_n}$$

is the  $n$ th convergent in the continued fraction expansion of  $\alpha$  then the Bruno-condition reads

$$\sum_n \frac{\log(q_{n+1})}{q_n} < \infty.$$

This condition turns out to be necessary and sufficient for  $\Phi$  having positive radius of convergence (Yoccoz [1995](#), [2002](#)).

### Cremer's Example in Herman's Version

As an example consider the map

$$F(z) = \lambda z + z^2,$$

where  $\lambda \in \mathbb{T}^1$  is not a root of unity.

Observe that a point  $z \in \mathbb{C}$  is a periodic point of  $F$  with period  $q$  if and only if  $F^q(z) = z$ , where obviously

$$F^q(z) = \lambda^q z + \dots + z^{2^q}.$$

Writing

$$F^q(z) - z = z(\lambda^q - 1 + \dots + z^{2^q-1}),$$

the period  $q$  periodic points exactly are the roots of the right hand side polynomial. Abbreviating  $N = 2^q - 1$ , it directly follows that, if  $z_1, z_2, \dots, z_N$  are the nontrivial roots, then for their product we have

$$z_1 \cdot z_2 \cdot \dots \cdot z_N = \lambda^q - 1.$$

It follows that there exists a nontrivial root within radius

$$|\lambda^q - 1|^{1/N}$$

of  $z = 0$ .

Now consider the set of  $\lambda \in \mathbb{T}^1$  defined as follows:  $\lambda \in \mathcal{A}$  whenever

$$\liminf_{q \rightarrow \infty} |\lambda^q - 1|^{1/N} = 0.$$

It can be directly shown that  $\mathcal{A}$  is *residual*, again compare with Oxtoby ([1971](#)). It also follows that for  $\lambda \in \mathcal{A}$  linearization is impossible. Indeed, since the rotation is irrational, the existence of periodic points in any neighborhood of  $z = 0$  implies zero radius of convergence.

### Remark

- Notice that the residual set  $\mathcal{A}$  is in the complement of the full measure set of all Diophantine numbers, again see Oxtoby ([1971](#)).
- Considering  $\lambda \in \mathbb{T}^1$  as a parameter, we see a certain analogy of these results on complex linearization with the Theorems [3](#) and [4](#). Indeed, in this case for a full measure set of  $\lambda$ s on a neighborhood of  $z = 0$  the map  $F = F_\lambda$  is conjugated to a rigid irrational rotation.

Such a domain in the  $z$ -plane often is referred to as a Siegel disc. For a more general discussion of these and of Herman rings, see Milnor ([2006](#)).

## KAM Theory: An Overview

In section “[Invariant Curves of Planar Diffeomorphisms](#)” we described the persistence of quasi-periodicity in the setting of diffeomorphisms of the circle or the plane. The general perturbation theory of quasi-periodic motions is known under the name Kolmogorov–Arnold–Moser (or KAM) theory and discussed extensively elsewhere in this encyclopedia ▶ “[Kolmogorov-Arnold-Moser \(KAM\) Theory for Finite and Infinite Dimensional Systems](#)”. Presently we briefly summarize parts of this KAM theory in broad terms, as this fits in our considerations, thereby largely referring to Arnold ([1963](#)), Kolmogorov ([1954](#), [1957](#)), Pöschel ([1982](#), [2001](#)), Zehnder ([1974](#), [1975](#)), also see Broer ([2004](#)), Broer and Sevryuk ([2007](#)), and Ciocci et al. ([2005](#)).

In general quasi-periodicity is defined by a smooth conjugation. First on the  $n$ -torus  $\mathbb{T}^n = \mathbb{R}^n / (2\pi\mathbb{Z})^n$  consider the vector field

$$\mathbb{X}_\omega = \sum_{j=1}^n \omega_j \frac{\partial}{\partial \varphi_j},$$

where  $\omega_1, \omega_2, \dots, \omega_n$  are called frequencies (Broer et al. 2007b; Moser 1967). Now, given a smooth (say, of class  $C^\infty$ ) vector field  $X$  on a manifold  $M$ , with  $T \subseteq M$  an invariant  $n$ -torus, we say that the restriction  $X|_T$  is *parallel* if there exists  $\omega \in \mathbb{R}^n$  and a smooth diffeomorphism  $\Phi : T \rightarrow \mathbb{T}^n$ , such that  $\Phi_*(X|_T) = \mathbb{X}_\omega$ . We say that  $X|_T$  is *quasi-periodic* if the frequencies  $\omega_1, \omega_2, \dots, \omega_n$  are independent over  $\mathbb{Q}$ .

A quasi-periodic vector field  $X|_T$  leads to an integer affine structure on the torus  $T$ . In fact, since each orbit is dense, it follows that the self conjugations of  $\mathbb{X}_\omega$  exactly are the translations of  $\mathbb{T}^n$ , which completely determine the affine structure of  $\mathbb{T}^n$ . Then, given  $\Phi : T \rightarrow \mathbb{T}^n$  with  $\Phi_*(X|_T) = \mathbb{X}_\omega$ , it follows that the self conjugations of  $X|_T$  determines a natural affine structure on the torus  $T$ . Note that the conjugation  $\Phi$  is unique modulo translations in  $T$  and  $\mathbb{T}^n$ .

Note that the composition of  $\Phi$  by a translation of  $\mathbb{T}^n$  does not change the frequency vector  $\omega$ . However, the composition by a linear invertible map  $S \in \text{GL}(n, \mathbb{Z})$  yields  $S_*\mathbb{X}_\omega = \mathbb{X}_{S\omega}$ . We here speak of an *integer affine* structure (Broer et al. 2007b).

### Remark

- The transition maps of an integer affine structure are translations and elements of  $\text{GL}(n, \mathbb{Z})$ .
- The current construction is compatible with the integrable affine structure on the Liouville tori of an integrable Hamiltonian system (Arnold 1978). Note that in that case the structure extends to all parallel tori.

### Classical KAM Theory

The classical KAM theory deals with smooth, nearly integrable Hamiltonian systems of the form

$$\begin{aligned}\dot{\varphi} &= \omega(I) + \varepsilon f(I, \varphi, \varepsilon) \\ \dot{I} &= \varepsilon g(I, \varphi, \varepsilon),\end{aligned}\tag{6}$$

where  $I$  varies over an open subset of  $\mathbb{R}^n$  and  $\varphi$  over the standard torus  $\mathbb{T}^n$ . Note that for  $\varepsilon = 0$  the phase space as an open subset of  $\mathbb{R}^n \times \mathbb{T}^n$  is foliated by invariant tori, parametrized by  $I$ .

Each of the tori is parametrized by  $\varphi$  and the corresponding motion is parallel (or multi-periodic or conditionally periodic) with frequency vector  $\omega(I)$ .

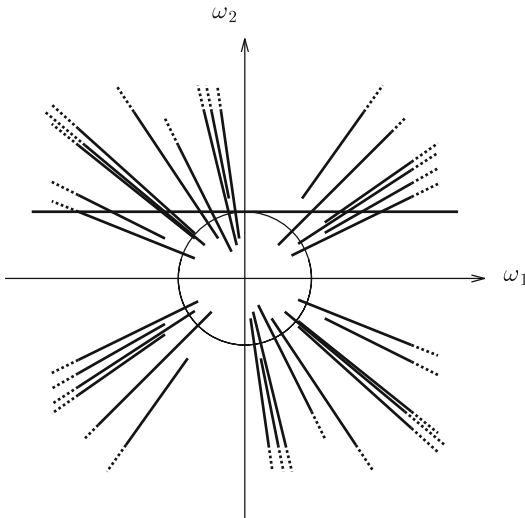
Perturbation theory asks for persistence of the invariant  $n$ -tori and the parallelity of their motion for small values of  $|\varepsilon|$ . The answer that KAM theory gives needs two essential ingredients. The first ingredient is that of *Kolmogorov non-degeneracy* which states that the map  $I \in \mathbb{R}^n \mapsto \omega(I) \in \mathbb{R}^n$  is a (local) diffeomorphism. Compare with the twist condition of section “Invariant Curves of Planar Diffeomorphisms”. The second ingredient generalizes the Diophantine conditions (5) of that section as follows: for  $\tau > n - 1$  and  $\gamma > 0$  consider the set

$$\mathbb{R}_{\tau, \gamma}^n = \left\{ \omega \in \mathbb{R}^n \mid \left| \langle \omega, k \rangle \right| \geq \gamma |k|^{-\tau}, k \in \mathbb{Z}^n \setminus \{0\} \right\}.\tag{7}$$

The following properties are more or less direct. First  $\mathbb{R}_{\tau, \gamma}^n$  has a closed half line geometry in the sense that if  $\omega \in \mathbb{R}_{\tau, \gamma}^n$  and  $s \geq 1$  then also  $s\omega \in \mathbb{R}_{\tau, \gamma}^n$ . Moreover, the intersection  $\mathbb{S}^{n-1} \cap \mathbb{R}_{\tau, \gamma}^n$  is a Cantor set of measure  $\mathbb{S}^{n-1} \setminus \mathbb{R}_{\tau, \gamma}^n = O(\gamma)$  as  $\gamma \downarrow 0$ , see Fig. 4.

Completely in the spirit of Theorem 4, the classical KAM theorem roughly states that a Kolmogorov non-degenerate nearly integrable system (6) <sub>$\varepsilon$</sub> , for  $|\varepsilon| \ll 1$  is smoothly conjugated to the unperturbed version (6)<sub>0</sub>, provided that the frequency map  $\omega$  is co-restricted to the Diophantine set  $\mathbb{R}_{\tau, \gamma}^n$ . In this formulation smoothness has to be taken in the sense of Whitney (Pöschel 1982; Wagener 2003, also compare with Broer 2004; Broer and Sevryuk 2007; Broer et al. 1990, 1996; Ciocci et al. 2005; Pöschel 2001).

As a consequence we may say that in Hamiltonian systems of  $n$  degrees of freedom typically quasi-periodic invariant (Lagrangian)  $n$ -tori occur with positive measure in phase space. It should be said that also an iso-energetic version of this classical result exists, implying a similar conclusion restricted to energy hypersurfaces (Arnold 1978; Arnold and Avez 1967; Broer and Huitema 1991; Broer and Sevryuk 2007). The Twist Map



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 4** The Diophantine set  $\mathbb{R}_{\tau,\gamma}^n$  has the closed half line geometry and the intersection  $\mathbb{S}^{n-1} \cap \mathbb{R}_{\tau,\gamma}^n$  is a Cantor set of measure  $\mathbb{S}^{n-1} \setminus \mathbb{R}_{\tau,\gamma}^n = O(\gamma)$  as  $\gamma \downarrow 0$

Theorem 4 is closely related to the iso-energetic KAM Theorem.

#### Remark

- We chose the quasi-periodic stability format as in section “Invariant Curves of Planar Diffeomorphisms”. For regularity issues compare with a remark following Theorem 3.
- For applications we largely refer to the introduction and to Broer and Sevryuk (2007), Broer et al. (1996), and references therein.
- Continuing the discussion on affine structures at the beginning of this section, we mention that by means of the symplectic form, the domain of the  $I$ -variables in  $\mathbb{R}^n$  inherits an affine structure (Duistermaat 1980), also see Lukina (2008) and references therein.

Statistical Mechanics deals with particle systems that are large, often infinitely large. The *Ergodic Hypothesis* roughly says that in a bounded energy hypersurface, the dynamics are ergodic, meaning that any evolution in the energy level set comes near every point of this set.

The taking of limits as the number of particles tends to infinity is a notoriously difficult subject. Here we discuss a few direct consequences of

classical KAM theory for many degrees of freedom. This discussion starts with Kolmogorov’s papers (1954, 1957), which we now present in a slightly rephrased form. First, we recall that for Hamiltonian systems (say, with  $n$  degrees of freedom), typically the union of Diophantine quasi-periodic Lagrangian invariant  $n$ -tori fills up positive measure in the phase space and also in the energy hypersurfaces. Second, such a collection of KAM tori immediately gives rise to non-ergodicity, since it clearly implies the existence of distinct invariant sets of positive measure. For background on Ergodic Theory, see e.g., Arnold and Avez (1967), Broer and Takens (2008), and Broer and Sevryuk (2007) for more references. Apparently the KAM tori form an obstruction to ergodicity, and a question is how bad this obstruction is as  $n \rightarrow \infty$ . Results in Arnold (1964) and de Jong (1999) indicate that this KAM theory obstruction is not too bad as the size of the system tends to infinity. In general the role of the Ergodic Hypothesis in Statistical Mechanics has turned out to be much more subtle than was expected, see e.g., Bricmont (1996) and Gallavotti et al. (2004).

#### Dissipative KAM Theory

As already noted by Moser (1966, 1967), KAM theory extends outside the world of Hamiltonian systems, like to volume preserving systems, or to equivariant or reversible systems. This also holds for the class of general smooth systems, often called dissipative. In fact, the KAM theorem allows for a Lie algebra proof, that can be used to cover all these special cases (Broer and Sevryuk 2007; Broer et al. 1990, 1996, 2007d). It turns out that in many cases parameters are needed for persistent occurrence of (Diophantine) quasi-periodic tori.

As an example we now consider the dissipative setting, where we discuss a parametrized system with normally hyperbolic invariant  $n$ -tori carrying quasi-periodic motion. From Hirsch et al. (1977) it follows that this is a persistent situation and that, up to a smooth (in this case of class  $C^k$  for large  $k$ ) diffeomorphism, we can restrict to the case where  $\mathbb{T}^n$  is the phase space. To fix thoughts we consider the smooth system

$$\begin{aligned}\dot{\varphi} &= \omega(\mu) + \varepsilon f(\varphi, \mu, \varepsilon) \\ \dot{\mu} &= 0\end{aligned}\quad (8)$$

where  $\mu \in \mathbb{R}^n$  is a multi-parameter. The results of the classical KAM theorem regarding (6) <sub>$\varepsilon$</sub>  largely carry over to (8) <sub>$\mu, \varepsilon$</sub> .

Now, for  $\varepsilon = 0$  the product of phase space and parameter space as an open subset of  $\mathbb{T}^n \times \mathbb{R}^n$  is completely foliated by invariant  $n$ -tori and since the perturbation does not concern the  $\dot{\mu}$ -equation, this foliation is persistent. The interest is with the dynamics on the resulting invariant tori that remains parallel after the perturbation; compare with the setting of Theorem 3. As just stated, KAM theory here gives a solution similar to the Hamiltonian case. The analogue of the Kolmogorov non-degeneracy condition here is that the frequency map  $\mu \mapsto \omega(\mu)$  is a (local) diffeomorphism. Then, in the spirit of Theorem 3, we state that the system (8) <sub>$\mu, \varepsilon$</sub>  is smoothly conjugated to (8) <sub>$\mu, 0$</sub> , as before, provided that the map  $\omega$  is co-restricted to the Diophantine set  $\mathbb{R}_{\tau, \gamma}^n$ . Again the smoothness has to be taken in the sense of Whitney (Broer et al. 1990; Pöschel 1982; Wagener 2003; Zehnder 1974, 1975; also see Broer 2004; Broer and Sevryuk 2007; Broer et al. 1996; Ciocci et al. 2005).

It follows that the occurrence of normally hyperbolic invariant tori carrying (Diophantine) quasi-periodic flow is typical for families of systems with sufficiently many parameters, where this occurrence has positive measure in parameter space. In fact, if the number of parameters equals the dimension of the tori, the geometry as sketched in Fig. 4 carries over in a diffeomorphic way.

### Remark

- Many remarks following subsection “Classical KAM Theory” and Theorem 3 also hold here.
- In cases where the system is degenerate, for instance because there is a lack of parameters, a path formalism can be invoked, where the parameter path is required to be a generic sub-family of the Diophantine set  $\mathbb{R}_{\tau, \gamma}^n$ , see Fig. 4. This amounts to the Rüssmann non-degeneracy, that still gives positive measure of quasi-periodicity in the parameter space, compare with Broer and Sevryuk (2007), Broer et al. (1996), and references therein.

- In the dissipative case the KAM theorem gives rise to families of quasi-periodic attractors in a typical way. This is of importance in center manifold reductions of infinite dimensional dynamics as, e.g., in fluid mechanics (Ruelle 1989; Ruelle and Takens 1971). In section “Transition to Chaos and Turbulence” we shall return to this subject.

### Lower Dimensional Tori

We extend the above approach to the case of lower dimensional tori, i.e., where the dynamics transversal to the tori is also taken into account. We largely follow the set-up of Broer et al. (1990, 2007d) that follows Moser (1967). Also see Broer and Sevryuk (2007), Broer et al. (1996), and references therein. Changing notation a little, we now consider the phase space  $\mathbb{T}^n \times \mathbb{R}^m = \{x(\text{mod } 2\pi), y\}$ , as well a parameter space  $\{\mu\} = P \subset \mathbb{R}^s$ . We consider a  $C^\infty$ -family of vector fields  $X(x, y, \mu)$  as before, having  $\mathbb{T}^n \times \{0\} \subset \mathbb{T}^n \times \mathbb{R}^m$  as an invariant  $n$ -torus for  $\mu = \mu_0 \in P$ .

$$\begin{aligned}\dot{x} &= \omega(\mu) + f(y, \mu) \\ \dot{y} &= \Omega(\mu)y + g(y, \mu) \\ \dot{\mu} &= 0\end{aligned}\quad (9)$$

with  $f(y, \mu_0) = O(|y|)$  and  $g(y, \mu_0) = O(|y|^2)$ , so we assume the invariant torus to be of Floquet type.

The system  $X = X(x, y, \mu)$  is integrable in the sense that it is  $\mathbb{T}^n$ -symmetric, i.e.,  $x$ -independent (Broer et al. 1990). The interest is with the fate of the invariant torus  $\mathbb{T}^n \times \{0\}$  and its parallel dynamics under small perturbation to a system  $\tilde{X} = \tilde{X}(x, y, \mu)$  that no longer needs to be integrable.

Consider the smooth mappings  $\omega : P \rightarrow \mathbb{R}^n$  and  $\Omega : P \rightarrow \text{gl}(m, \mathbb{R})$ . To begin with we restrict to the case where all eigenvalues of  $\Omega(\mu_0)$  are simple and nonzero. In general for such a matrix  $\Omega \in \text{gl}(m, \mathbb{R})$ , let the eigenvalues be given by  $\alpha_1 \pm i\beta_1, \dots, \alpha_{N_1} \pm i\beta_{N_1}$  and  $\delta_1, \dots, \delta_{N_2}$ , where all  $\alpha_j, \beta_j$  and  $\delta_j$  are real and hence  $m = 2N_1 + N_2$ . Also consider the map  $\text{spec} : \text{gl}(m, \mathbb{R}) \rightarrow \mathbb{R}^{2N_1+N_2}$ , given by  $\Omega \mapsto (\alpha, \beta, \delta)$ . Next to the internal frequency vector  $\omega \in \mathbb{R}^n$ , we also have the vector  $\beta \in \mathbb{R}^{N_1}$  of normal frequencies.



The present analogue of Kolmogorov non-degeneracy is the Broer–Huiteima–Takens (BHT) non-degeneracy condition (Broer et al. 1990; Sevryuk 2007), which requires that the product map  $\omega \times (\text{spec}) \circ \Omega : P \rightarrow \mathbb{R}^n \times \text{gl}(m, \mathbb{R})$  at  $\mu = \mu_0$  has a surjective derivative and hence is a local submersion (Hirsch 1976).

Furthermore, we need Diophantine conditions on both the internal and the normal frequencies, generalizing Eq. (7). Given  $\tau > n - 1$  and  $\gamma > 0$ , it is required for all  $k \in \mathbb{Z}^n \setminus \{0\}$  and all  $\ell \in \mathbb{Z}^{N_1}$  with  $|\ell| \leq 2$  that

$$|\langle k, \omega \rangle + \langle \ell, \beta \rangle| \geq \gamma |k|^{-\tau}. \quad (10)$$

Inside  $\mathbb{R}^n \times \mathbb{R}^{N_1} = \{\omega, \beta\}$  this yields a Cantor set as before (compare Fig. 4). This set has to be pulled back along the submersion  $\omega \times (\text{spec}) \circ \Omega$ , for examples see subsections “[\(n-1\)-Tori](#)” and “[Quasi-periodic Bifurcations](#)” below.

The KAM theorem for this setting is quasi-periodic stability of the  $n$ -tori under consideration, as in subsection “[Dissipative KAM Theory](#)”, yielding typical examples where quasi-periodicity has positive measure in parameter space. In fact, we get a little more here, since the normal linear behavior of the  $n$ -tori is preserved by the Whitney smooth conjugations. This is expressed as normal linear stability, which is of importance for quasi-periodic bifurcations, see subsection “[Quasi-periodic Bifurcations](#)” below.

#### Remark

- A more general set-up of the normal stability theory (Broer et al. 2007d) adapts the above to the case of non-simple (multiple) eigenvalues. Here the BHT non-degeneracy condition is formulated in terms of versal unfolding of the matrix  $\Omega(\mu_0)$  (Arnold 1983). For possible conditions under which vanishing eigenvalues are admissible see Broer et al. (1990, 2009), Hanßmann (2007), and references therein.
- This general set-up allows for a structure preserving formulation as mentioned earlier, thereby including the Hamiltonian and volume preserving case, as well as equivariant and reversible cases. This allows us, for example, to deal with quasi-periodic versions of the

Hamiltonian and the reversible Hopf bifurcation (Broer et al. 2006a, 2007a, c, 2009).

- The Parameterized KAM Theory discussed here a priori needs many parameters. In many cases the parameters are distinguished in the sense that they are given by action variables, etc. For an example see subsection “[\(n-1\)-Tori](#)” on Hamiltonian  $(n-1)$ -tori. Also see Sevryuk (2007), Broer and Sevryuk (2007), and Broer et al. (1996) where the case of Rüssmann non-degeneracy is included. This generalizes a remark at the end of subsection “[Dissipative KAM Theory](#)”.

#### Global KAM Theory

We stay in the Hamiltonian setting, considering Lagrangian invariant  $n$ -tori as these occur in a Liouville integrable system with  $n$  degrees of freedom. The union of these tori forms a smooth  $\mathbb{T}^n$ -bundle  $f : M \rightarrow B$  (where we leave out all singular fibers). It is known that this bundle can be non-trivial (Cushman and Bates 1997; Duistermaat 1980) as can be measured by monodromy and Chern class. In this case global action angle variables are not defined. This non-triviality, among other things, is of importance for semi-classical versions of the classical system at hand, in particular for certain spectrum defects (Cushman et al. 2004; Efsthafiou 2005; Vũ Ngọc 1999; Waalkens et al. 2003), for more references also see Broer and Sevryuk (2007).

Restricting to the classical case, the problem is what happens to the (non-trivial)  $\mathbb{T}^n$ -bundle  $f$  under small, non-integrable perturbation. From the classical KAM theory, see subsection “[Classical KAM Theory](#)” we already know that on trivializing charts of  $f$  Diophantine quasi-periodic  $n$ -tori persist. In fact, at this level, a Whitney smooth conjugation exists between the integrable system and its perturbation, which is even Gevrey regular (Wagener 2003). It turns out that these local KAM conjugations can be glued together so to obtain a global conjugation at the level of quasi-periodic tori, thereby implying global quasi-periodic stability (Broer et al. 2007b). Here we need unicity of KAM tori, i.e., independence of the action-angle chart used in the classical KAM theorem (Broer and Takens 2007). The proof uses the integer

affine structure on the quasi-periodic tori, which enables taking convex combinations of the local conjugations subjected to a suitable partition of unity (Hirsch 1976; Spivak 1970). In this way the geometry of the integrable bundle can be carried over to the nearly-integrable one.

The classical example of a Liouville integrable system with non-trivial monodromy (Cushman and Bates 1997; Duistermaat 1980) is the spherical pendulum, which we now briefly revisit. The configuration space is  $\mathbb{S}^2 = \{q \in \mathbb{R}^3 \mid \langle q, q \rangle = 1\}$  and the phase space  $T^*\mathbb{S}^2 \cong \{(q, p) \in \mathbb{R}^6 \mid \langle q, q \rangle = 1 \text{ and } \langle q, p \rangle = 0\}$ . The two integrals  $I = q_1 p_2 - q_2 p_1$  (angular momentum) and  $E = \frac{1}{2} \langle p, p \rangle + q_3$  (energy) lead to an energy momentum map  $\mathcal{EM} : T^*\mathbb{S}^2 \rightarrow \mathbb{R}^2$ , given by  $(q, p) \mapsto (I, E) = (q_1 p_2 - q_2 p_1, \frac{1}{2} \langle p, p \rangle + q_3)$ . In Fig. 5 we show the image of the map  $\mathcal{EM}$ . The shaded area  $B$  consists of regular values, the fiber above which is a Lagrangian two-torus; the union of these gives rise to a bundle  $f : M \rightarrow B$  as described before, where  $f = \mathcal{EM}|_M$ . The motion in the two-tori is a superposition of Huygens' rotations and pendulum-like swinging, and the non-existence of global action angle variables reflects that the three interpretations of 'rotating oscillation', 'oscillating rotation' and 'rotating rotation' cannot be reconciled in a consistent way. The singularities of the fibration include the equilibria  $(q, p) = ((0, 0, \pm 1), (0, 0, 0)) \mapsto (I, E) = (0, \pm 1)$ . The boundary of

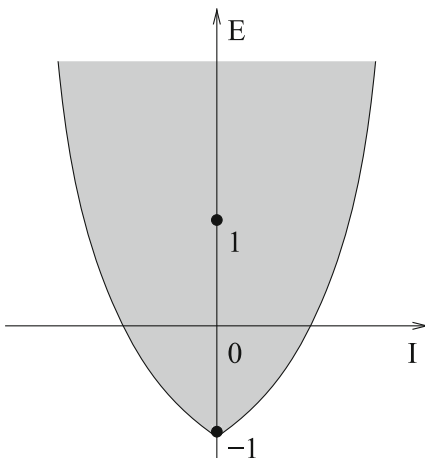
this image also consists of singular points, where the fiber is a circle that corresponds to Huygens' horizontal rotations of the pendulum. The fiber above the upper equilibrium point  $(I, E) = (0, 1)$  is a pinched torus (Cushman and Bates 1997), leading to non-trivial monodromy, in a suitable bases of the period lattices, given by

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \in \text{GL}(2, \mathbb{R}).$$

The question here is what remains of the bundle  $f$  when the system is perturbed. Here we observe that locally Kolmogorov non-degeneracy is implied by the non-trivial monodromy (Zung 1996; Rink 2004). From Broer et al. (2007b) and Rink (2004) it follows that the non-trivial monodromy can be extended in the perturbed case.

#### Remark

- The case where this perturbation remains integrable is covered in Matveev (1996), but presently the interest is with the nearly integrable case, so where the axial symmetry is broken. Also compare Broer and Sevryuk (2007) and many of its references.
- The global conjugations of Broer et al. (2007b) are Whitney smooth (even Gevrey regular (Wagener 2003)) and near the identity map in the  $C^\infty$ -topology (Hirsch 1976). Geometrically speaking these diffeomorphisms also are  $\mathbb{T}^n$ -bundle isomorphisms between the unperturbed and the perturbed bundle, the basis of which is a Cantor set of positive measure.



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 5** Range of the energy-momentum map of the spherical pendulum

### Splitting of Separatrices

KAM theory does not predict the fate of close-to-resonant tori under perturbations. For fully resonant tori the phenomenon of frequencylocking leads to the destruction of the torus under (sufficiently rich) perturbations, and other resonant tori disintegrate as well. In the case of a single resonance between otherwise Diophantine frequencies the perturbation leads to quasi-periodic bifurcations, cf. section “[Transition to Chaos and Turbulence](#)”.



While KAM theory concerns the fate of most trajectories and for all times, a complementary theorem has been obtained in Marco and Sauzin (2003), Nekhoroshev (1977, 1985), and Niederman (2004). It concerns all trajectories and states that they stay close to the unperturbed tori for *long* times that are exponential in the inverse of the perturbation strength. For trajectories starting close to surviving tori the diffusion is even superexponentially slow, cf. Morbidelli and Giorgilli (1995a, b). Here a form of smoothness exceeding the mere existence of infinitely many derivatives of the Hamiltonian is a necessary ingredient, for finitely differentiable Hamiltonians one only obtains polynomial times.

Solenoids, which cannot be present in integrable systems, are constructed for generic Hamiltonian systems in Birkhoff (1935), Markus and Meyer (1980), and Meiss (1986), yielding the simultaneous existence of representatives of all homeomorphism-classes of solenoids. Hyperbolic tori form the core of a construction proposed in Arnold (1964) of trajectories that venture off to distant points of the phase space. In the unperturbed system the union of a family of hyperbolic tori, parametrized by the actions conjugate to the toral angles, form a normally hyperbolic manifold. The latter is persistent under perturbations, cf. Hirsch et al. (1977) and Mielke (1991), and carries a Hamiltonian flow with fewer degrees of freedom. The main difference between integrable and non-integrable systems already occurs for periodic orbits.

### Periodic Orbits

A sharp difference to dissipative systems is that it is generic for hyperbolic periodic orbits on compact energy shells in Hamiltonian systems to have homoclinic orbits, cf. Abraham and Marsden (1978) and references therein. For integrable systems these form together a pinched torus, but under generic perturbations the stable and unstable manifold of a hyperbolic periodic orbit intersect transversely. It is a nontrivial task to actually check this genericity condition for a given non-integrable perturbation, a first-order condition going back to Poincaré requires the computation of the so-called Mel'nikov integral, see Guckenheimer and Holmes (1983) and Wiggins (1990) for more details. In two degrees of freedom normalization leads to

approximations that are integrable to all orders, which implies that the Melnikov integral is a flat function. In the real analytic case the Melnikov criterion is still decisive in many examples (Gelfreich and Lazutkin 2001).

Genericity conditions are traditionally formulated in the universe of smooth vector fields, and this makes the whole class of analytic vector fields appear to be non-generic. This is an overly pessimistic view as the conditions defining a certain class of generic vector fields may certainly be satisfied by a given analytic system. In this respect it is interesting that the generic properties may also be formulated in the universe of analytic vector fields, see Broer and Tangerman (1986) for more details.

### ( $n-1$ )-Tori

The  $(n-1)$ -parameter families of invariant  $(n-1)$ -tori organize the dynamics of an integrable Hamiltonian system in  $n$  degrees of freedom, and under small perturbations the parameter space of persisting analytic tori is Cantorized. This still allows for a global understanding of a substantial part of the dynamics, but also leads to additional questions.

A hyperbolic invariant torus  $\mathbb{T}^{n-1}$  has its Floquet exponents off the imaginary axis. Note that  $\mathbb{T}^{n-1}$  is not a normally hyperbolic manifold. Indeed, the normal linear behavior involves the  $n-1$  zero eigenvalues in the direction of the parametrizing actions as well; similar to Eq. (9) the format

$$\begin{aligned}\dot{x} &= \omega(y) + O(y) + O(z^2) \\ \dot{y} &= O(y) + O(z^3) \\ \dot{z} &= \Omega(y)z + O(z^2)\end{aligned}$$

in Floquet coordinates yields an  $x$ -independent matrix  $\Omega$  that describes the symplectic normal linear behavior, cf. Broer et al. (1990). The union  $\{z=0\}$  over the family of  $(n-1)$ -tori is a normally hyperbolic manifold and constitutes the center manifold of  $\mathbb{T}^{n-1}$ . Separatrices splitting yields the dividing surfaces in the sense of Wiggins et al. (2001).

The persistence of elliptic tori under perturbation from an integrable system involves not only

the internal frequencies of  $\mathbb{T}^{n-1}$ , but also the normal frequencies. Next to the internal resonances the necessary Diophantine conditions (10) exclude the normal-internal resonances

$$\langle k, \omega \rangle = \alpha_j \quad (11)$$

$$\langle k, \omega \rangle = 2\alpha_j \quad (12)$$

$$\langle k, \omega \rangle = \alpha_i + \alpha_j \quad (13)$$

$$\langle k, \omega \rangle = \alpha_i - \alpha_j. \quad (14)$$

The first three resonances lead to the quasi-periodic center-saddle bifurcation studied in section “[Transition to Chaos and Turbulence](#)”, the frequency-halving (or quasi-periodic period doubling) bifurcation and the quasi-periodic Hamiltonian Hopf bifurcation, respectively. The resonance (14) generalizes an equilibrium in 1 : 1 resonance whence  $\mathbb{T}^{n-1}$  persists and remains elliptic, cf. de Jong (1999). When passing through resonances (12) and (13) the lower-dimensional tori lose ellipticity and acquire hyperbolic Floquet exponents. Elliptic  $(n - 1)$ -tori have a single normal frequency whence Eqs. (11) and (12) are the only normal-internal resonances. See Broer et al. (2003b) for a thorough treatment of the ensuing possibilities.

The restriction to a single normal-internal resonance is dictated by our present possibilities. Indeed, already the bifurcation of equilibria with a fourfold zero eigenvalue leads to unfoldings that simultaneously contain all possible normal resonances. Thus, a satisfactory study of such tori which already may form one-parameter families in integrable Hamiltonian systems with five degrees of freedom has to await further progress in local bifurcation theory.

## Transition to Chaos and Turbulence

One of the main interests over the second half of the twentieth century has been the transition between orderly and complicated forms of dynamics upon variation of either initial states or of system parameters. By ‘orderly’ we here mean equilibrium and periodic dynamics and

by complicated quasi-periodic and chaotic dynamics, although we note that only chaotic dynamics is associated to unpredictability, e.g. see Broer and Takens (2008). As already discussed in the introduction systems like a forced nonlinear oscillator or the planar three-body problem exhibit coexistence of periodic, quasi-periodic and chaotic dynamics, also compare with Fig. 1.

Similar remarks go for the onset of turbulence in fluid dynamics. Around 1950 this led to the scenario of Hopf–Landau–Lifschitz (Hopf 1942, 1948; Landau 1944; Landau and Lifschitz 1959), which roughly amounts to the following. Stationary fluid motion corresponds to an equilibrium point in an  $\infty$ -dimensional state space of velocity fields. The first transition is a Hopf bifurcation (Guckenheimer and Holmes 1983; Hopf 1942; Kuznetsov 2004), where a periodic solution branches off. In a second transition of similar nature a quasi-periodic two-torus branches off, then a quasi-periodic three-torus, etc. The idea is that the motion picks up more and more frequencies and thus obtains an increasingly complicated power spectrum. In the early 1970s this idea was modified in the Ruelle–Takens route to turbulence, based on the observation that, for flows, a three-torus can carry chaotic (or strange) attractors (Newhouse et al. 1978; Ruelle and Takens 1971), giving rise to a broad band power spectrum. By the quasi-periodic bifurcation theory (Broer and Sevryuk 2007; Broer et al. 1990, 1996) as sketched below these two approaches are unified in a generic way, keeping track of measure theoretic aspects. For general background in dynamical systems theory we refer to Broer and Takens (2008) and Katok and Hasselblatt (1995).

Another transition to chaos was detected in the quadratic family of interval maps

$$f_\mu(x) = \mu x(1 - x),$$

see Devaney (1989), de Melo and van Strien (1991), and Milnor (2006), also for a holomorphic version. This transition consists of an infinite sequence of period doubling bifurcations ending up in chaos; it has several universal aspects and occurs persistently in families of dynamical systems. In many of these cases also homoclinic bifurcations show up, where sometimes the transition to

chaos is immediate when parameters cross a certain boundary, for general theory see Benedicks and Carleson (1985, 1991), Broer et al. (1991), and Palis and Takens (1993). There exist quite a number of case studies where all three of the above scenarios play a role, e.g., see Broer et al. (1998, 2002, 2008a) and many of their references.

### Quasi-periodic Bifurcations

For the classical bifurcations of equilibria and periodic orbits, the bifurcation sets and diagrams are generally determined by a classical geometry in the product of phase space and parameter space as already established by, e.g., Arnold (1994) and Thom (1989), often using singularity theory. Quasi-periodic bifurcation theory concerns the extension of these bifurcations to invariant tori in nearly-integrable systems, e.g., when the tori lose their normal hyperbolicity or when certain (strong) resonances occur. In that case the dense set of resonances, also responsible for the small divisors, leads to a Cantorization of the classical geometries obtained from Singularity Theory (Broer et al. 1990, 2003b, 2005, 2006a, b, 2007a, c, d; in preparation; Chenciner 1985a; Hanßmann 1988, 2004, 2007, also see Broer and Sevryuk 2007; Broer et al. 1996; Chenciner and Iooss 1979; Ciocci et al. 2005). Broadly speaking, one could say that in these cases the Preparation Theorem (Thom 1989) is partly replaced by KAM theory. Since the KAM theory has been developed in several settings with or without preservation of structure, see section “KAM Theory: An Overview”, for the ensuing quasi-periodic bifurcation theory the same holds.

### Hamiltonian Cases

To fix thoughts we start with an example in the Hamiltonian setting, where a robust model for the quasi-periodic center-saddle bifurcation is given by

$$\begin{aligned} H_{\omega_1, \omega_2, \mu, \varepsilon}(I, \varphi, p, q) &= \omega_1 I_1 + \omega_2 I_2 \\ &+ \frac{1}{2} p^2 + V_\mu(q) \\ &+ \varepsilon f(I, \varphi, p, q) \end{aligned} \quad (15)$$

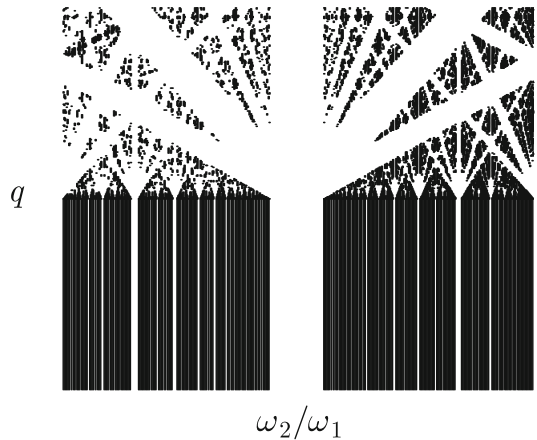
with  $V_\mu(q) = \frac{1}{3} q^3 - \mu q$ , compare with Hanßmann (1988, 2007). The unperturbed (or integrable)

case  $\varepsilon = 0$ , by factoring out the  $\mathbb{T}^2$ -symmetry, boils down to a standard center-saddle bifurcation, involving the fold catastrophe (Thom 1989) in the potential function  $V = V_\mu(q)$ . This results in the existence of two invariant two-tori, one elliptic and the other hyperbolic. For  $0 \neq |\varepsilon| \ll 1$  the dense set of resonances complicates this scenario, as sketched in Fig. 6, determined by the Diophantine conditions

$$\begin{aligned} |\langle k, \omega \rangle| &\geq \gamma |k|^{-\tau}, & \text{for } q < 0 \\ |\langle k, \omega \rangle + \ell \beta(q)| &\geq \gamma |k|^{-\tau}, & \text{for } q > 0 \end{aligned} \quad (16)$$

for all  $k \in \mathbb{Z}^n \setminus \{0\}$  and for all  $\ell \in \mathbb{Z}$  with  $|\ell| \leq 2$ . Here  $\beta(q) = \sqrt{2q}$  is the normal frequency of the elliptic torus given by  $q = \sqrt{\mu}$  for  $\mu > 0$ . As before, (cf. sections “Invariant Curves of Planar Diffeomorphisms” and “KAM Theory: An Overview”), this gives a Cantor set of positive measure (Broer and Sevryuk 2007; Broer et al. 1990, 1996, 2007d; Hanßmann 2007; Moser 1966, 1967).

For  $0 < |\varepsilon| \ll 1$  Fig. 6 will be distorted by a near-identity diffeomorphism; compare with the formulations of the Theorems 3 and 4. On the Diophantine Cantor set the dynamics is quasi-periodic, while in the gaps generically there is



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 6** Sketch of the Cantorized Fold, as the bifurcation set of the quasi-periodic center-saddle bifurcation for  $n = 2$  (Hanßmann 1988), where the horizontal axis indicates the frequency ratio  $\omega_2 : \omega_1$ , cf. (15). The lower part of the figure corresponds to hyperbolic tori and the upper part to elliptic ones. See the text for further interpretations

coexistence of periodicity and chaos, roughly comparable with Fig. 1, at left. The gaps at the border furthermore lead to the phenomenon of parabolic resonance, cf. Litvak-Hinenzon and Rom-Kedar (2002).

Similar programs exist for all cuspid and umbilic catastrophes (Broer et al. 2005, 2006b; Hanßmann 2004) as well as for the Hamiltonian Hopf bifurcation (Broer et al. 2006a, 2007c). For applications of this approach see Broer et al. (2003b). For a reversible analogue see Broer et al. (2007a). As so often within the gaps generically there is an infinite regress of smaller gaps (Baensens et al. 1991; Broer et al. 2003b). For theoretical background we refer to Broer et al. (1990, 2007d), Moser (1967), for more references also see Broer and Sevryuk (2007).

### Dissipative Cases

In the general dissipative case we basically follow the same strategy. Given the standard bifurcations of equilibria and periodic orbits, we get more complex situations when invariant tori are involved as well. The simplest examples are the quasi-periodic saddle-node and quasi-periodic period doubling (Broer et al. 1990; also see Broer and Sevryuk 2007; Broer et al. 1996).

To illustrate the whole approach let us start from the Hopf bifurcation of an equilibrium point of a vector field (Guckenheimer and Holmes 1983; Hopf 1942; Kuznetsov 2004; Palis and de Melo 1982) where a hyperbolic point attractor loses stability and branches off a periodic solution, cf. subsection “Dissipative Perturbations”. A topological normal form is given by

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - (y_1^2 + y_2^2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (17)$$

where  $y = (y_1, y_2) \in \mathbb{R}^2$ , ranging near  $(0, 0)$ . In this representation usually one fixes  $\beta = 1$  and lets  $\alpha = \mu$  (near 0) serve as a (bifurcation) parameter, classifying modulo topological equivalence. In polar coordinates (17) so gets the form

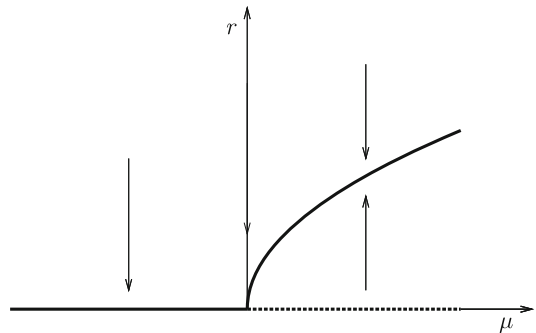
$$\begin{aligned} \dot{\varphi} &= 1, \\ \dot{r} &= \mu r - r^3. \end{aligned}$$

Figure 7 shows an amplitude response diagram (often called the bifurcation diagram). Observe the occurrence of the attracting periodic solution for  $\mu > 0$  of amplitude  $\sqrt{\mu}$ .

Let us briefly consider the Hopf bifurcation for fixed points of diffeomorphisms. A simple example has the form

$$P(y) = e^{2\pi(\alpha+i\beta)y} + O(|y|^2), \quad (18)$$

$y \in \mathbb{C} \cong \mathbb{R}^2$ , near 0. To start with  $\beta$  is considered a constant, such that  $\beta$  is not rational with denominator less than five, see Arnold (1983), Takens (1974b), and where  $O(|y|^2)$  should contain generic third order terms. As before, we let  $\alpha = \mu$  serve as a bifurcation parameter, varying near 0. On one side of the bifurcation value  $\mu = 0$ , this system has by normal hyperbolicity and Hirsch et al. (1977), an invariant circle. Here, due to the invariance of the rotation numbers of the invariant circles, no topological stability can be obtained (Newhouse et al. 1983). Still this bifurcation can be characterized by many persistent properties. Indeed, in a generic two-parameter family (18), say with both  $\alpha$  and  $\beta$  as parameters, the periodicity in the parameter plane is organized in resonance tongues (Arnold 1983; Broer et al. 2003a; Kuznetsov 2004). (The tongue structure is hardly visible when only one parameter, like  $\alpha$ , is used.) If the



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 7** Bifurcation diagram of the Hopf bifurcation

diffeomorphism is the return map of a periodic orbit for flows, this bifurcation produces an invariant two-torus. Usually this counterpart for flows is called Neïmark–Sacker bifurcation. The periodicity as it occurs in the resonance tongues, for the vector field is related to phase lock. The tongues are contained in gaps of a Cantor set of quasi-periodic tori with Diophantine frequencies. Compare the discussion in subsection “Circle Maps”, in particular also regarding the Arnold family and Fig. 3. Also see section “KAM Theory: An Overview” and again compare with Oxtoby (1971).

Quasi-periodic versions exist for the saddle-node, the period doubling and the Hopf bifurcation. Returning to the setting with  $\mathbb{T}^n \times \mathbb{R}^m$  as the phase space, we remark that the quasi-periodic saddle-node and period doubling already occur for  $m = 1$ , or in an analogous center manifold. The quasi-periodic Hopf bifurcation needs  $m \geq 2$ . We shall illustrate our results on the latter of these cases, compare with Broer (2003) and Broer et al. (1996). For earlier results in this direction see Chenciner and Iooss (1979). Our phase space is  $\mathbb{T}^n \times \mathbb{R}^2 = \{x(\bmod 2\pi), y\}$ , where we are dealing with the parallel invariant torus  $\mathbb{T}^n \times \{0\}$ . In the integrable case, by  $\mathbb{T}^n$ -symmetry we can reduce to  $\mathbb{R}^2 = \{y\}$  and consider the bifurcations of relative equilibria. The present interest is with small non-integrable perturbations of such integrable models.

We now discuss the *quasi-periodic Hopf bifurcation* (Braaksma and Broer 1987; Broer et al. 1990), largely following (Ciocci et al. 2005). The unperturbed, integrable family  $X = X_\mu(x, y)$  on  $\mathbb{T}^n \times \mathbb{R}^2$  has the form

$$X_\mu(x, y) = [\omega(\mu) + f(y, \mu)]\partial_x + [\Omega(\mu)y + g(y, \mu)]\partial_y, \quad (19)$$

where  $f = O(|y|)$  and  $g = O(|y|^2)$  as before. Moreover  $\mu \in P$  is a multi-parameter and  $\omega : P \rightarrow \mathbb{R}^n$  and  $\Omega : P \rightarrow \mathfrak{gl}(2, \mathbb{R})$  are smooth maps. Here we take

$$\Omega(\mu) = \begin{pmatrix} \alpha(\mu) & -\beta(\mu) \\ \beta(\mu) & \alpha(\mu) \end{pmatrix},$$

which makes the  $\partial_y$  component of Eq. (19) compatible with the planar Hopf family (17). The

present form of Kolmogorov non-degeneracy is Broer–Huiteima–Takens stability (Broer et al. 1990, 2009; Broer et al. 2007d), requiring that there is a subset  $\Gamma \subseteq P$  on which the map

$$\mu \in P \mapsto (\omega(\mu), \Omega(\mu)) \in \mathbb{R}^n \times \mathfrak{gl}(2, \mathbb{R})$$

is a submersion. For simplicity we even assume that  $\mu$  is replaced by

$$(\omega, (\alpha, \beta)) \in \mathbb{R}^n \times \mathbb{R}^2.$$

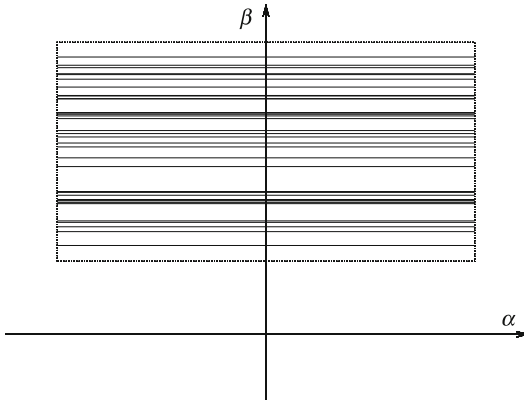
Observe that if the non-linearity  $g$  satisfies the well-known Hopf non-degeneracy conditions, e.g., compare (Guckenheimer and Holmes 1983; Kuznetsov 2004), then the relative equilibrium  $y = 0$  undergoes a standard planar Hopf bifurcation as described before. Here  $\alpha$  again plays the role of bifurcation parameter and a closed orbit branches off at  $\alpha = 0$ . To fix thoughts we assume that  $y = 0$  is attracting for  $\alpha < 0$ , and that the closed orbit occurs for  $\alpha > 0$ , and is attracting as well. For the integrable family  $X$ , qualitatively we have to multiply this planar scenario with  $\mathbb{T}^n$ , by which all equilibria turn into invariant attracting or repelling  $n$ -tori and the periodic attractor into an attracting invariant  $(n + 1)$ -torus. Presently the question is what happens to both the  $n$ - and the  $(n + 1)$ -tori, when we apply a small near-integrable perturbation.

The story runs much like before. Apart from the BHT non-degeneracy condition we require Diophantine conditions (10), defining the Cantor set

$$\begin{aligned} \Gamma_{\tau, \gamma}^{(2)} = \{(\omega, \alpha\beta) \in \Gamma \mid \\ \langle k, \omega \rangle + \ell\beta \geq \gamma|k|^{-\tau}, \forall k \in \mathbb{Z}^n \setminus \{0\}, \forall \ell \in \mathbb{Z} \text{ with } |\ell| \leq 2\}, \end{aligned} \quad (20)$$

In Fig. 8 we sketch the intersection of  $\Gamma_{\tau, \gamma}^{(2)} \subset \mathbb{R}^n \times \mathbb{R}^2$  with a plane  $\{\omega\} \times \mathbb{R}^2$  for a Diophantine (internal) frequency vector  $\omega$ , cf. Eq. (7).

From Braaksma and Broer (1987) and Broer et al. (1990) it now follows that for any family  $\tilde{X}$  on  $\mathbb{T}^n \times \mathbb{R}^2 \times P$ , sufficiently near  $X$  in the  $C^\infty$ -topology a near-identity  $C^\infty$ -diffeomorphism  $\Phi : \mathbb{T}^n \times \mathbb{R}^2 \times \Gamma \rightarrow \mathbb{T}^n \times \mathbb{R}^2 \times \Gamma$  exists, defined

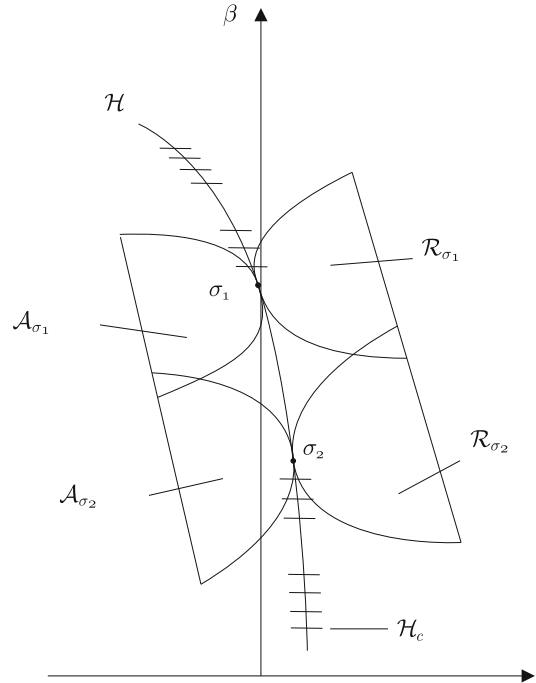


**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 8** Planar section of the Cantor set  $\Gamma_{\tau,\gamma}^{(2)}$

near  $\mathbb{T}^n \times \{0\} \times \Gamma$ , that conjugates  $X$  to  $\tilde{X}$  when further restricting to  $\mathbb{T}^n \times \{0\} \times \Gamma_{\tau,\gamma}^{(2)}$ . So this means that the Diophantine quasi-periodic invariant  $n$ -tori are persistent on a diffeomorphic image of the Cantor set  $\Gamma_{\tau,\gamma}^{(2)}$ , compare with the formulations of the Theorems 3 and 4.

Similarly we can find invariant  $(n+1)$ -tori. We first have to develop a  $\mathbb{T}^{n+1}$  symmetric normal form approximation (Braaksma and Broer 1987; Broer et al. 1990) and ► “Normal Forms in Perturbation Theory”. For this purpose we extend the Diophantine conditions (20) by requiring that the inequality holds for all  $|\ell| \leq N$  for  $N=7$ . We thus find another large Cantor set, again see Fig. 8, where Diophantine quasi-periodic invariant  $(n+1)$ -tori are persistent. Here we have to restrict to  $\alpha > 0$  for our choice of the sign of the normal form coefficient, compare with Fig. 7.

In both the cases of  $n$ -tori and of  $(n+1)$ -tori, the nowhere dense subset of the parameter space containing the tori can be fattened by normal hyperbolicity to open subsets. Indeed, the quasi-periodic  $n$ - and  $(n+1)$ -tori are infinitely normally hyperbolic (Hirsch et al. 1977). Exploiting the normal form theory (Braaksma and Broer 1987; Broer et al. 1990) and ► “Normal Forms in Perturbation Theory” to the utmost and using a more or less standard contraction argument (Braaksma and Broer 1987; Chow and Hale 1982), a fattening of the parameter domain with invariant tori can be obtained that leaves out only small ‘bubbles’ around the resonances, as sketched and



**Hamiltonian Perturbation Theory (and Transition to Chaos), Fig. 9** Fattening by normal hyperbolicity of a nowhere dense parameter set with invariant  $n$ -tori in the perturbed system. The curve  $\mathcal{H}$  is the Whitney smooth (even Gevrey regular (Wagener 2003)) image of the  $\beta$ -axis in Fig. 8.  $\mathcal{H}$  interpolates the Cantor set  $\mathcal{H}_c$  that contains the non-hyperbolic Diophantine quasi-periodic invariant  $n$ -tori, corresponding to  $\Gamma_{\tau,\gamma}^{(2)}$ , see Eq. (20). To the points  $\sigma_{1,2} \in \mathcal{H}_c$  discs  $\mathcal{A}_{\sigma_{1,2}}$  are attached where we find attracting normally hyperbolic  $n$ -tori and similarly in the discs  $\mathcal{R}_{\sigma_{1,2}}$  repelling ones. The contact between the disc boundaries and  $\mathcal{H}$  is infinitely flat (Braaksma and Broer 1987; Broer et al. 1990)

explained in Fig. 9 for the  $n$ -tori. For earlier results in the same spirit in a case study of the quasi-periodic saddle-node bifurcation see Chenciner (1985a, b, 1988), also compare with Baensens et al. (1991).

### A Scenario for the Onset of Turbulence

Generally speaking, in many settings quasi-periodicity constitutes the order in between chaos (Broer et al. 1996). In the Hopf–Landau–Lifschitz–Ruelle–Takens scenario (Hopf 1948; Landau 1944; Landau and Lifschitz 1959; Ruelle and Takens 1971) we may consider a sequence of typical transitions as given by quasi-periodic Hopf bifurcations, starting with the standard



Hopf or Hopf–Neĭmark–Sacker bifurcation as described before. In the gaps of the Diophantine Cantor sets generically there will be coexistence of periodicity, quasi-periodicity and chaos in infinite regress. As said earlier, period doubling sequences and homoclinic bifurcations may accompany this.

As an example consider a family of maps that undergoes a generic quasi-periodic Hopf bifurcation from circle to two-torus. It turns out that here the Cantorized fold of Fig. 6 is relevant, where now the vertical coordinate is a bifurcation parameter. Moreover compare with Fig. 3, where also variation of  $\varepsilon$  is taken into account. The Cantor set contains the quasi-periodic dynamics, while in the gaps we can have chaos, e.g., in the form of Hénon like strange attractors (Broer et al. 2008a; Newhouse et al. 1978). A fattening process as explained above, also can be carried out here.

## Future Directions

One important general issue is the mathematical characterization of chaos and ergodicity in dynamical systems, in conservative, dissipative and in other settings. This is a tough problem as can already be seen when considering two-dimensional diffeomorphisms. In particular we refer to the still unproven ergodicity conjecture of Arnold and Avez (1967) and to the conjectures around Hénon like attractors and the principle ‘Hénon everywhere’, compare with Broer and Krauskopf (2000) and Broer et al. (1998). For a discussion see subsection “[A Scenario for the Onset of Turbulence](#)”. In higher dimension this problem is even harder to handle, e.g., compare with Broer et al. (2008a, b), and references therein. In the conservative case a related problem concerns a better understanding of Arnold diffusion.

Somewhat related to this is the analysis of dynamical systems without an explicit perturbation setting. Here numerical and symbolic tools are expected to become useful to develop computer assisted proofs in extended perturbation settings, diagrams of Lyapunov exponents, symbolic dynamics, etc. Compare with Simó (2001). Also see Broer et al. (2008a, b) for applications and

further reference. This part of the theory is important for understanding concrete models, that often are not given in perturbation format.

Regarding nearly-integrable Hamiltonian systems, several problems have to be considered. Continuing the above line of thought, one interesting development is the development of Hamiltonian bifurcation theory without integrable normal form and, likewise, of KAM theory without action angle coordinates (de la Llave et al. 2005). One big related issue also is to develop KAM theory outside the perturbation format.

The previous section addressed persistence of Diophantine tori involved in a bifurcation. Similar to Cremer’s example in subsection “[Cremer’s Example in Herman’s Version](#)” the dynamics in the gaps between persistent tori displays new phenomena. A first step has been made in Litvak-Hinenzon and Rom-Kedar (2002) where internally resonant parabolic tori involved in a quasi-periodic Hamiltonian pitchfork bifurcation are considered. The resulting large dynamical instabilities may be further amplified for tangent (or flat) parabolic resonances, which fail to satisfy the iso-energetic non-degeneracy condition.

The construction of solenoids in Birkhoff (1935) and Markus and Meyer (1980) uses elliptic periodic orbits as starting points, the simplest example being the result of a period-doubling sequence. This construction should carry over to elliptic tori, where normal-internal resonances lead to encircling tori of the same dimension, while internal resonances lead to elliptic tori of smaller dimension and excitation of normal modes increases the torus dimension. In this way one might be able to construct solenoid-type invariant sets that are limits of tori with varying dimension.

Concerning the global theory of nearly-integrable torus bundles (Broer et al. 2007b), it is of interest to understand the effects of quasi-periodic bifurcations on the geometry and its invariants. Also it is of interest to extend the results of Vũ Ngọc (1999) when passing to semiclassical approximations. In that case two small parameters play a role, namely Planck’s constant as well as the distance away from integrability.

## Bibliography

- Abraham R, Marsden JE (1978) Foundations of mechanics, 2nd edn. Benjamin, London
- Arnol'd VI, Kozlov VV, Neishtadt AI (1988) Mathematical aspects of classical and celestial mechanics. In: Arnold VI (ed) Dynamical systems, vol III. Springer, New York
- Arnold VI (1961) Small divisors I: on mappings of the circle onto itself. *Izv Akad Nauk SSSR Ser Mat* 25: 21–86. (in Russian); English translation: *Am Math Soc Transl Ser* 2(46):213–284 (1965); Erratum: *Izv Akad Nauk SSSR Ser Mat* 28:479–480 (1964, in Russian)
- Arnold VI (1962) On the classical perturbation theory and the stability problem of the planetary system. *Dokl Akad Nauk SSSR* 145:487–490
- Arnold VI (1963) Proof of a theorem by A.N. Kolmogorov on the persistence of conditionally periodic motions under a small change of the Hamilton function. *Russ Math Surv* 18(5):9–36. (English; Russian original)
- Arnold VI (1964) Instability of dynamical systems with several degrees of freedom. *Sov Math Dokl* 5:581–585
- Arnold VI (1978) Mathematical methods of classical mechanics, GTM 60. Springer, New York
- Arnold VI (1983) Geometrical methods in the theory of ordinary differential equations. Springer, New York
- Arnold VI (1994) Dynamical systems V: bifurcation theory and catastrophe theory. In: *Encyclopedia of mathematical sciences*, vol 5. Springer, New York
- Arnold VI, Avez A (1967) *Problèmes Ergodiques de la Mécanique classique*. Gauthier-Villars, Paris. English edition: Arnold VI, Avez A (1968) *Ergodic problems of classical mechanics*. Benjamin, New York
- Baesens C, Guckenheimer J, Kim S, MacKay RS (1991) Three coupled oscillators: mode-locking, global bifurcation and toroidal chaos. *Phys D* 49(3):387–475
- Barrow-Green J (1997) Poincaré and the three body problem. In: *History of mathematics*, vol 11. American Mathematical Society/London Mathematical Society, Providence/London
- Benedicks M, Carleson L (1985) On iterations of  $1 - ax^2$  on  $(-1, 1)$ . *Ann Math* 122:1–25
- Benedicks M, Carleson L (1991) The dynamics of the Hénon map. *Ann Math* 133:73–169
- Benettin G (2005) Physical applications of Nekhoroshev theorem and exponential estimates. In: Giorgilli A (ed) *Hamiltonian dynamics theory and applications*, Cetraro 1999, *Lecture notes in mathematics*, vol 1861. Springer, New York, pp 1–76
- Birkhoff BD (1935) Nouvelles recherches sur les systèmes dynamiques. *Mem Pont Acad Sci Novi Lyncae* 1(3): 85–216
- Braaksma BLJ, Broer HW (1987) On a quasi-periodic Hopf bifurcation. *Ann Inst Henri Poincaré* 4(2): 115–168
- Bricmont J (1996) Science of chaos or chaos in science? In: Gross PR, Levitt N, Lewis MW (eds) *The flight from science and reason*. New York Academy of Sciences, New York, pp 131–175. (New York, 1995; *Annals of the New York Academy of Sciences*, vol 775. Also appeared in: *Phys Mag* 17:159–208 (1995))
- Broer HW (2003) Coupled Hopf-bifurcations: persistent examples of  $n$ -quasi periodicity determined by families of 3-jets. *Astérisque* 286:223–229
- Broer HW (2004) KAM theory: the legacy of Kolmogorov's 1954 paper. *Bull Am Math Soc* 41(4):507–521
- Broer HW, Huitema GB (1991) A proof of the isoenergetic KAM-theorem from the “ordinary” one. *J Differ Equ* 90:52–60
- Broer HW, Krauskopf B (2000) Chaos in periodically driven systems. In: Krauskopf B, Lenstra D (eds) *Fundamental issues of nonlinear laser dynamics*, AIP conference proceedings 548. American Institute of Physics, Melville, pp 31–53
- Broer HW, Roussarie R (2001) Exponential confinement of chaos in the bifurcation set of real analytic diffeomorphisms. In: Broer HW, Krauskopf B, Vegter G (eds) *Global analysis of dynamical systems*, Festschrift dedicated to Floris Takens for his 60th birthday. IOP, Bristol/Philadelphia, pp 167–210
- Broer HW, Sevryuk MB (2007) KAM theory: quasi-periodicity in dynamical systems. In: Broer HW, Hasselblatt B, Takens F (eds) *Handbook of dynamical systems*, vol 3. North-Holland, Amsterdam. (to appear)
- Broer HW, Takens F (1989) Formally symmetric normal forms and genericity. *Dyn Rep* 2:36–60
- Broer HW, Takens F (2007) Unicity of KAM tori. *Ergod Theory Dyn Syst* 27:713–724
- Broer HW, Takens F (2008) *Dynamical systems and chaos*. To be Published by Epsilon Uitgaven, Amsterdam
- Broer HW, Tangerman FM (1986) From a differentiable to a real analytic perturbation theory, applications to the Kupka Smale theorems. *Ergod Theory Dyn Syst* 6: 345–362
- Broer HW, Huitema GB, Takens F, Braaksma BLJ (1990) Unfoldings and bifurcations of quasi-periodic tori. In: *Memoir AMS*, vol 421. American Mathematical Society, Providence
- Broer HW, Dumortier F, van Strien SJ, Takens F (1991) Structures in dynamics, finite dimensional deterministic studies. In: de Jager EM, van Groesen EWC (eds) *Studies in mathematical physics*, vol II. North-Holland, Amsterdam
- Broer HW, Huitema GB, Sevryuk MB (1996) Quasi-periodic motions in families of dynamical systems: order amidst chaos, *Lecture notes in mathematics*, vol 1645. Springer, New York
- Broer HW, Simó C, Tatjer JC (1998) Towards global models near homoclinic tangencies of dissipative diffeomorphisms. *Nonlinearity* 11(3):667–770
- Broer HW, Simó C, Vitolo R (2002) Bifurcations and strange attractors in the Lorenz-84 climate model with seasonal forcing. *Nonlinearity* 15(4):1205–1267
- Broer HW, Golubitsky M, Vegter G (2003a) The geometry of resonance tongues: a singularity theory approach. *Nonlinearity* 16:1511–1538
- Broer HW, Hanßmann H, Jorba À, Villanueva J, Wagener FOO (2003b) Normal-internal resonances in quasi-



- periodically forced oscillators: a conservative approach. *Nonlinearity* 16:1751–1791
- Broer HW, Hoveijn I, Lunter G, Vegter G (2003c) Bifurcations in Hamiltonian systems: computing singularities by Gröbner bases, Lecture notes in mathematics, vol 1806. Springer, New York
- Broer HW, Hanßmann H, You J (2005) Bifurcations of normally parabolic tori in Hamiltonian systems. *Nonlinearity* 18:1735–1769
- Broer HW, Hanßmann H, Hoo J, Naudot V (2006a) Nearly-integrable perturbations of the Lagrange top: applications of KAM theory. In: Denteneer D, den Hollander F, Verbitskiy E (eds) *Dynamics & stochasticity: festschrift in honor of MS Keane*, Lecture notes, vol 48. Institute of Mathematical Statistics, Beachwood, pp 286–303
- Broer HW, Hanßmann H, You J (2006b) Umbilical torus bifurcations in Hamiltonian systems. *J Differ Equ* 222: 233–262
- Broer HW, Naudot V, Roussarie R (2006c) Catastrophe theory in Dulacun foldings. *Ergod Theory Dyn Syst* 26: 1–35
- Broer HW, Ciocci MC, Hanßmann H (2007a) The quasi-periodic reversible Hopf bifurcation. In: Doedel E, Krauskopf B, Sanders J (eds) *Recent advances in nonlinear dynamics: theme section dedicated to André Vanderbauwhede*. *Intern J Bifurc Chaos* 17:2605–2623
- Broer HW, Cushman RH, Fassò F, Takens F (2007b) Geometry of KAM tori for nearly integrable Hamiltonian systems. *Ergod Theory Dyn Syst* 27(3):725–741
- Broer HW, Hanßmann H, Hoo J (2007c) The quasi-periodic Hamiltonian Hopf bifurcation. *Nonlinearity* 20:417–460
- Broer HW, Hoo J, Naudot V (2007d) Normal linear stability of quasi-periodic tori. *J Differ Equ* 232:355–418
- Broer HW, Simó C, Vitolo R (2008a) The Hopf–Saddle-Node bifurcation for fixed points of 3D-diffeomorphisms, the Arnol’d resonance web. *Bull Belg Math Soc Simon Stevin* 15:769–787
- Broer HW, Simó C, Vitolo R (2008b) The Hopf–Saddle-Node bifurcation for fixed points of 3D-diffeomorphisms, analysis of a resonance “bubble”. *Phys D Nonlinear Phenom* (to appear)
- Broer HW, Ciocci MC, Hanßmann H, Vanderbauwhede A (2009) Quasi-periodic stability of normally resonant tori. *Phys D* 238:309–318
- Broer HW, Hanßmann H, You J (in preparation) On the destruction of resonant Lagrangean tori in Hamiltonian systems
- Chenciner A (1985a) Bifurcations de points fixes elliptiques – I, courbes invariantes. *Publ Math IHÉS* 61:67–127
- Chenciner A (1985b) Bifurcations de points fixes elliptiques – II, orbites périodiques et ensembles de Cantor invariants. *Invent Math* 80:81–106
- Chenciner A (1988) Bifurcations de points fixes elliptiques III, orbites périodiques de “petites” périodes et élimination résonnante des couples de courbes invariantes. *Publ Math IHÉS* 66:5–91
- Chenciner A, Iooss G (1979) Bifurcations de tores invariants. *Arch Ration Mech Anal* 69(2):109–198; 71(4): 301–306
- Chow S-N, Hale JK (1982) *Methods of bifurcation theory*. Springer, New York
- Chow S-N, Li C, Wang D (1994) *Normal forms and bifurcation of planar vector fields*. Cambridge University Press, Cambridge, UK
- Ciocci MC, Litvak-Hinenzon A, Broer HW (2005) Survey on dissipative KAM theory including quasi-periodic bifurcation theory based on lectures by Henk Broer. In: Montaldi J, Ratiu T (eds) *Geometric mechanics and symmetry: the Peyresq lectures, LMS lecture notes series*, vol 306. Cambridge University Press, Cambridge, UK, pp 303–355
- Cushman RH, Bates LM (1997) *Global aspects of classical integrable systems*. Birkhäuser, Basel
- Cushman RH, Dullin HR, Giacobbe A, Holm DD, Joyeux M, Lynch P, Sadovskii DA, Zhilinskii BI (2004) CO<sub>2</sub> molecule as a quantum realization of the 1:1:2 resonant swing-spring with monodromy. *Phys Rev Lett* 93:024302
- de Jong HH (1999) Quasiperiodic breathers in systems of weakly coupled pendulums: applications of KAM theory to classical and statistical mechanics. Ph.D. thesis, University of Groningen, Groningen
- de la Llave R, González A, Jorba À, Villanueva J (2005) KAM theory without action-angle variables. *Nonlinearity* 18:855–895
- de Melo W, van Strien SJ (1991) *One-dimensional dynamics*. Springer, New York
- Devaney RL (1989) *An introduction to chaotic dynamical systems*, 2nd edn. Addison-Wesley, Redwood City
- Diacu F, Holmes P (1996) *Celestial encounters. The origins of chaos and stability*. Princeton University Press, Princeton
- Duistermaat JJ (1980) On global action-angle coordinates. *Commun Pure Appl Math* 33:687–706
- Dumortier F, Roussarie R, Sotomayor J (1991) Generic 3-parameter families of vector fields, unfoldings of saddle, focus and elliptic singularities with nilpotent linear parts. In: Dumortier F, Roussarie R, Sotomayor J, Zoladek H (eds) *Bifurcations of planar vector fields: nilpotent singularities and Abelian integrals*, LNM 1480. Springer, New York, pp 1–164
- Efstathiou K (2005) *Metamorphoses of Hamiltonian systems with symmetries*, LNM, vol 1864. Springer, Heidelberg
- Féjoz J (2004) Démonstration du “théorème d’Arnold” sur la stabilité du système planétaire (d’après Herman). *Ergod Theory Dyn Syst* 24:1–62
- Gallavotti G, Bonetto F, Gentile G (2004) *Aspects of ergodic, qualitative and statistical theory of motion*. Springer, New York
- Gelfreich VG, Lazutkin VF (2001) Splitting of separatrices: perturbation theory and exponential smallness. *Russ Math Surv* 56:499–558
- Guckenheimer J, Holmes P (1983) *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer, New York

- Hanßmann H (1988) The quasi-periodic centre-saddle bifurcation. *J Differ Equ* 142:305–370
- Hanßmann H (2004) Hamiltonian torus bifurcations related to simple singularities. In: Ladde GS, Medhin NG, Sambandham M (eds) *Dynamic systems and applications*. Dynamic Publishers, Atlanta, pp 679–685
- Hanßmann H (2007) Local and semi-local bifurcations in Hamiltonian dynamical systems – results and examples, *Lecture notes in mathematics*, vol 1893. Springer, Berlin
- Herman M (1977) Mesure de Lebesgue et nombre de rotation. In: Palis J, do Carmo M (eds) *Geometry and topology*, *Lecture notes in mathematics*, vol 597. Springer, Berlin, pp 271–293
- Herman MR (1979) Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publ Math IHÉS* 49:5–233
- Hirsch MW (1976) *Differential topology*. Springer, Berlin
- Hirsch MW, Pugh CC, Shub M (1977) *Invariant manifolds*, *Lecture notes in mathematics*, vol 583. Springer, Berlin
- Hofer H, Zehnder E (1994) *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser, Basel
- Hopf E (1942) Abzweigung einer periodischen Lösung von einer stationären Lösung eines differential systems. *Ber Math Phys Kl Sächs Akad Wiss Leipzig* 94:1–22
- Hopf E (1948) A mathematical example displaying features of turbulence. *Commun Appl Math* 1:303–322
- Huygens C (1888–1950) *Œuvres complètes de Christiaan Huygens*, vol 5, pp 241–263; vol 17, pp 156–189. Martinus Nijhoff, The Hague
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*. Cambridge University Press, Cambridge, UK
- Kolmogorov AN (1954) On the persistence of conditionally periodic motions under a small change of the Hamilton function. *Dokl Akad Nauk SSSR* 98:527–530. (in Russian). English translation: Stochastic behavior in classical and quantum Hamiltonian systems, *Volta memorial conference* (Como, 1977). In: Casati G, Ford J (eds) *Lecture notes in physics*, vol 93. Springer, Berlin pp 51–56 (1979). Reprinted in: Bai Lin Hao (ed) *Chaos*. World Scientific, Singapore, pp 81–86 (1984)
- Kolmogorov AN (1957) The general theory of dynamical systems and classical mechanics. In: Gerretsen JCH, de Groot J (eds) *Proceedings of the International Congress of Mathematicians*, vol 1 (1954). North-Holland, Amsterdam, pp 315–333. (in Russian). Reprinted in: *International Mathematical Congress in Amsterdam* (1954) (Plenary lectures). Fizmatgiz, Moscow, pp 187–208 (1961). English translation as Appendix D in: Abraham RH (1967) *Foundations of mechanics*. Benjamin, New York, pp 263–279. Reprinted as Appendix in [1], pp 741–757
- Kuznetsov YA (2004) *Elements of applied bifurcation theory*, *Applied mathematical sciences*, vol 112, 3rd edn. Springer, New York
- Landau LD (1944) On the problem of turbulence. *Akad Nauk* 44:339
- Landau LD, Lifschitz EM (1959) *Fluid mechanics*. Pergamon, Oxford, UK
- Laskar J (1995) Large scale chaos and marginal stability in the Solar System. In: Iagolnitzer D (ed) *XIth International Congress of Mathematical Physics* (Paris, 1994). Internat Press, Cambridge, UK, pp 75–120
- Litvak-Hinenzon A, Rom-Kedar V (2002) Parabolic resonances in 3 degree of freedom near-integrable Hamiltonian systems. *Phys D* 164:213–250
- Lochak P (1999) Arnold diffusion; a compendium of remarks and questions. In: Simó C (ed) *Hamiltonian systems with three or more degrees of freedom* (S’Agaró, 1995), *NATO ASI series C: math phys sci*, vol 533. Kluwer, Dordrecht, pp 168–183
- Lochak P, Marco J-P (2005) Diffusion times and stability exponents for nearly integrable analytic systems. *Cent Eur J Math* 3:342–397
- Lochak P, Neishtadt AI (1992) Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian. *Chaos* 2:495–499
- Lukina O (2008) *Geometry of torus bundles in Hamiltonian systems*. Ph.D. thesis, University of Groningen, Groningen
- MacKay RS (1993) *Renormalisation in area-preserving maps*. World Scientific, Singapore
- Marco J-P, Sauzin D (2003) Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems. *Publ Math Inst Hautes Etud Sci* 96:199–275
- Markus L, Meyer KR (1980) Periodic orbits and solenoids in generic Hamiltonian dynamical systems. *Am J Math* 102:25–92
- Matveev VS (1996) Integrable Hamiltonian systems with two degrees of freedom. Topological structure of saturated neighborhoods of points of focus-focus and saddle-saddle types. *Sb Math* 187:495–524
- McDuff D, Salamon D (1995) *Introduction to symplectic geometry*. Oxford University Press, Clarendon
- Meiss JD (1986) Class renormalization: islands around islands. *Phys Rev A* 34:2375–2383
- Meyer KR, Hall GR (1992) *Introduction to Hamiltonian dynamical systems and the N-body problem*, *Applied mathematical sciences*, vol 90. Springer, New York
- Mielke A (1991) *Hamiltonian and Lagrangian flows on center manifolds – with applications to elliptic variational problems*, *Lecture notes in mathematics*, vol 1489. Springer, New York
- Milnor JW (2006) *Dynamics in one complex variable*, *Annals of mathematical studies*, vol 160, 3rd edn. Princeton University Press, Princeton
- Morbidelli A, Giorgilli A (1995a) Superexponential stability of KAM tori. *J Stat Phys* 78:1607–1617
- Morbidelli A, Giorgilli A (1995b) On a connection between KAM and Nekhoroshev’s theorems. *Phys D* 86:514–516
- Moser JK (1962) On invariant curves of area-preserving mappings of an annulus. *Nachr Akad Wiss Göttingen II Math Phys Kl* 1:1–20
- Moser JK (1966) On the theory of quasiperiodic motions. *SIAM Rev* 8(2):145–172

- Moser JK (1967) Convergent series expansions for quasi-periodic motions. *Math Ann* 169:136–176
- Moser JK (1968) Lectures on Hamiltonian systems. *Mem Am Math Soc* 81:1–60
- Moser JK (1973) Stable and random motions in dynamical systems, with specialempphasis to celestial mechanics, *Annals of mathematical studies*, vol 77. Princeton University Press, Princeton
- Nekhoroshev NN (1977) An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems. *Russ Math Surv* 32:1–65
- Nekhoroshev NN (1985) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems II. In: Oleinik OA (ed) *Topics in modern mathematics*, Petrovskii seminar no. 5. Consultants Bureau, New York, pp 1–58
- Newhouse SE, Ruelle D, Takens F (1978) Occurrence of strange AxiomA attractors near quasi-periodic flows on  $T^m$ ,  $m \geq 3$ . *Commun Math Phys* 64:35–40
- Newhouse SE, Palis J, Takens F (1983) Bifurcations and stability of families of diffeomorphisms. *Publ Math IHÉS* 57:5–71
- Niederman L (2004) Prevalence of exponential stability amongnearly-integrable Hamiltonian systems. *Ergod Theory Dyn Syst* 24(2):593–608
- Oxtoby J (1971) *Measure and category*. Springer, New York
- Palis J, de Melo M (1982) *Geometric theory of dynamical systems*. Springer, New York
- Palis J, Takens F (1993) *Hyperbolicity & sensitive chaotic dynamics at homoclinic bifurcations*. Cambridge University Press, Cambridge, UK
- Poincaré H (1980) Sur le problème des trois corps et les équations de ladinamique. *Acta Math* 13:1–270
- Pöschel J (1982) Integrability of Hamiltonian systems on Cantor sets. *Commun Pure Appl Math* 35(5):653–696
- Pöschel J (1993) Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math Z* 213:187–216
- Pöschel J (2001) A lecture on the classical KAM theorem. *Proc Symp Pure Math* 69:707–732
- Rink BW (2004) A Cantor set of tori with monodromy near a focus-focus singularity. *Nonlinearity* 17:347–356
- Robinson C (1995) *Dynamical systems*. CRC Press, Boca Raton
- Roussarie R (1997) Smoothness properties of bifurcation diagrams. *Publ Mat* 41:243–268
- Ruelle D (1989) *Elements of differentiable dynamics and bifurcation theory*. Academic Press, Cambridge, MA
- Ruelle D, Takens F (1971) On the nature of turbulence. *Commun Math Phys* 20:167–192; 23:343–344
- Sevryuk MB (2007) Invariant tori in quasi-periodic non--autonomous dynamical systems via Herman's method. *Discrete Contin Dyn Syst A* 18(2/3):569–595
- Simó C (2001) Global dynamics and fast indicators. In: Broer HW, Krauskopf B, Vegter G (eds) *Global analysis of dynamical systems*, festschrift dedicated to Floris Takens for his 60th birthday. IOP, Bristol/Philadelphia, pp 373–390
- Spivak M (1970) *Differential geometry*, vol I. Publish or Perish, Boston
- Takens F (1973) Introduction to global analysis. *Comm. 2 of the Math. Inst. Rijksuniversiteit Utrecht*
- Takens F (1974a) Singularities of vector fields. *Publ Math IHÉS* 43:47–100
- Takens F (1974b) Forced oscillations and bifurcations. In: *Applications of global analysis I*, Comm 3 of the Math Inst Rijksuniversiteit Utrecht. (In: Broer HW, Krauskopf B, Vegter G (eds) *Global analysis of dynamical systems*, festschrift dedicated to Floris Takens for his 60th birthday. IOP, Bristol/Philadelphia, pp 1–62)
- Thom R (1989) *Structural stability and morphogenesis*. an outline of a general theory of models, 2nd edn. Addison-Wesley, Redwood City. (English; French original)
- Vũ Ngọc S (1999) Quantum monodromy in integrable systems. *Commun Math Phys* 203:465–479
- Waalkens H, Junge A, Dullin HR (2003) Quantum monodromy in the two-centre problem. *J Phys A Math Gen* 36:L307–L314
- Wagener FOO (2003) A note on Gevrey regular KAM theory and the inverse approximation lemma. *Dyn Syst* 18:159–163
- Wiggins S (1990) *Introduction to applied nonlinear dynamical systems and chaos*. Springer, New York
- Wiggins S, Wiesenfeld L, Jaffe C, Uzer T (2001) Impenetrable barriers in phase-space. *Phys Rev Lett* 86(24): 5478–5481
- Yoccoz J-C (1983)  $C^1$ -conjugaisons des difféomorphismes du cercle. In: Palis J (ed) *Geometric dynamics*, proceedings, Rio de Janeiro (1981). *Lecture notes in mathematics*, vol 1007, pp 814–827
- Yoccoz J-C (1992) Travaux de Herman sur les tores invariants. In: *Séminaire Bourbaki*, vol 754, 1991–1992. *Astérisque* 206:311–344
- Yoccoz J-C (1995) Théorème de Siegel, nombres de Bruno et polynômes quadratiques. *Astérisque* 231:3–88
- Yoccoz J-C (2002) Analytic linearization of circle diffeomorphisms. In: Marmi S, Yoccoz J-C (eds) *Dynamical systems and small divisors*, *Lecture notes in mathematics*, vol 1784. Springer, New York, pp 125–174
- Zehnder E (1974) An implicit function theorem for small divisor problems. *Bull Am Math Soc* 80(1):174–179
- Zehnder E (1975) Generalized implicit function theorems with applications to some small divisor problems, I and II. *Commun Pure Appl Math* 28(1):91–140; 29(1): 49–111 (1976)
- Zung NT (1996) Kolmogorov condition for integrable systems with focus-focus singularities. *Phys Lett A* 215(1/2):40–44

# Perturbation Theory in Quantum Mechanics

Luigi E. Picasso, Luciano Bracci and  
Emilio d'Emilio  
Dipartimento di Fisica, Università di Pisa, Pisa,  
Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Presentation of the Problem and an Example  
Perturbation of Point Spectra:  
    Nondegenerate Case  
Perturbation of Point Spectra: Degenerate Case  
The Brillouin-Wigner Method  
Symmetry and Degeneracy  
Problems with the Perturbation Series  
Perturbation of the Continuous Spectrum  
Time-Dependent Perturbations  
Future Directions  
Bibliography

## Glossary

**Hilbert Space** A Hilbert space  $\mathcal{H}$  is a normed complex vector space with a Hermitian scalar product. If  $\varphi, \psi \in \mathcal{H}$ , the scalar product between  $\varphi$  and  $\psi$  is written as  $(\varphi, \psi) \equiv (\psi, \varphi)^*$  and is taken to be linear in  $\psi$  and antilinear in  $\varphi$ : if  $a, b \in \mathbb{C}$ , the scalar product between  $a\varphi$  and  $b\psi$  is  $a^*b(\varphi, \psi)$ . The norm of  $\psi$  is defined as  $\|\psi\| \equiv \sqrt{(\psi, \psi)}$ . With respect to the norm,  $\|\cdot\|$ ,  $\mathcal{H}$  is a complete metric space. In the following,  $\mathcal{H}$  will be assumed to be separable, that is, any complete orthonormal set of vectors is countable.

**States and Observables** In quantum mechanics the states of a system are represented as vectors in a Hilbert space  $\mathcal{H}$ , with the convention that proportional vectors represent the same state. Physicists mostly use Dirac's notation: the elements of  $\mathcal{H}$  are represented by  $|\cdot\rangle$  ("ket"), and the scalar product between  $|\varphi\rangle$  and  $|\psi\rangle$  is written as  $\langle\varphi|\psi\rangle$  ("braket"). The observables, i.e., the physical quantities that can be measured, are represented by linear Hermitian (more precisely: self-adjoint) operators on  $\mathcal{H}$ . The eigenvalues of an observable are the only possible results of the measurement of the observable. The observables of a system are generally the same with the corresponding classical system – energy, angular momentum, etc., i.e., they are of the form  $f(q, p)$ , with  $q \equiv (q_1, \dots, q_n)$ ,  $p \equiv (p_1, \dots, p_n)$ ; the position and momentum canonical variables of the system ( $q_i$  and  $p_i$ ) are observables, i.e. operators, which satisfy the commutation relations  $[q_i, q_j] \equiv q_i q_j - q_j q_i = 0$ ,  $[p_i, p_j] = 0$ ,  $[q_i, p_j] = i\hbar\delta_{ij}$ , with  $\hbar$  the Planck constant  $h$  divided by  $2\pi$ .

**Representations** Since Separable Hilbert spaces are isomorphic, it is always possible to represent the elements of  $\mathcal{H}$  as elements of  $l_2$ , the space of the sequences  $\{u_i\}$ ,  $u_i \in \mathbb{C}$ , with the scalar product  $(v, u) \equiv \sum_i v_i^* u_i$ . This can be done by choosing an orthonormal basis of vectors  $e_i$  in  $\mathcal{H}$ :  $(e_i, e_j) = \delta_{ij}$  and defining  $u_i = (e_i, u)$ , with Dirac's notations  $|A\rangle \rightarrow \{a_i\}$ ,  $a_i = \langle e_i | A \rangle$ . Linear operators  $\xi$  are then represented by  $\{\xi_{ij}\}$ ,  $\xi_{ij} = (e_i, \xi e_j) \equiv \langle e_i | \xi | e_j \rangle$ . The  $\xi_{ij}$  are called "matrix elements" of  $\xi$  in the representation  $e_j$ . If  $\xi^\dagger$  is the Hermitian conjugate of  $\xi$ , then  $(\xi^\dagger)_{ij} = \xi_{ji}^*$ . If the  $e_i$  are eigenvectors of  $\xi$ , then the (infinite) matrix  $\xi_{ij}$  is diagonal, the diagonal elements being the eigenvalues of  $\xi$ .

**Schrödinger Equation** Among the observables, the Hamiltonian  $H$  plays a special role. It determines the time evolution of the system through the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi,$$

and its eigenvalues are the energy levels of the system. The eigenvalue equation  $H\psi = E\psi$  is called the Schrödinger equation.

**Schrödinger Representation** A different possibility is to represent the elements of  $\mathcal{H}$  as elements of  $L^2[\mathbb{R}^n]$ , the space of the square-integrable functions on  $\mathbb{R}^n$ , where  $n$  is the number of degrees of freedom of the system. This can be done by assigning how the operators  $q_i$  and  $p_i$  act on the functions of  $L^2[\mathbb{R}^n]$ : in the Schrödinger representation, the  $q_i$  are taken to act as multiplication by  $x_i$  and the  $p_i$  as  $-i\hbar \partial / \partial x_i$ ; if  $|A\rangle \rightarrow \psi_A(x_1, \dots, x_n)$ , then

$$\begin{aligned} q_i |A\rangle &\rightarrow x_i \psi_A(x_1, \dots, x_n), p_i |A\rangle \\ &\rightarrow -i\hbar \partial \psi_A(x_1, \dots, x_n) / \partial x_i. \end{aligned}$$

## Definition of the Subject

In the investigation of natural phenomena, a crucial role is played by the comparison between theoretical predictions and experimental data. Those practicing the two arts of the trade continuously put challenges to one another either presenting data which ask for an explanation or proposing new experimental verifications of a theory. Celestial mechanics offers the first historical instance of this interplay: the elliptical planetary orbits discovered by Kepler were explained by Newton; when discrepancies from the elliptical paths definitely emerged, it was necessary to add the effects of the heavier planets to the dominant role of the sun, until persistent discrepancies between theory and experiment asked for the drastic revision of the theory of gravitation put forth by Einstein, a revision which in turn offered a lot of new effects to observe, some of which have been verified only recently.

In this dialectic interaction between theory and experiment, only the simplest problem, that of a planet moving in the field of the sun within Newton's theory, can be solved exactly. All the rest was calculated by means of perturbation theory. Generally speaking, perturbation theory is the

technique of finding an approximate solution to a problem where to a dominant factor, which allows for an exact solution (zeroth-order solution), other “perturbing” factors are added which are outweighed by the dominant factor and are expected to bring small corrections to the zeroth order solution.

Perturbation theory is ever pervasive in physics, but an area where it plays a major role is quantum mechanics. In the early days of this discipline, the interpretation of atomic spectra was made possible only by a heavy use of perturbation theory, since the only exactly soluble problem was that of the hydrogen atom without external fields. The explanation of the Stark spectra (hydrogen in a constant electric field) and of the Zeeman spectra (atom in a magnetic field) was only possible when a perturbation theory tailored to the Schrödinger equation, which rules the atomic world, was devised. As for heavier atoms, in no case an exact solution for the Schrödinger equation is available: they could only be treated as a perturbation of simpler “hydrogenoid” atoms. Most of the essential aspects of atomic and molecular physics could be explained quantitatively in a few years by recourse to suitable forms of perturbation theory. Not only did it explain the position of the spectral lines, but also their relative intensities, and the absence of some lines which showed the impossibility of the corresponding transitions (selection rules), found a convincing explanation when symmetry considerations were introduced. When later more accurate measurements revealed details in the hydrogen spectrum (the Lamb shift) that only quantum field theory was able to explain, perturbation theory gained a new impetus which sometimes resulted in the anticipation of theory (quantum electrodynamics) over experiment as to the accuracy of the effect to be measured.

An attempt to describe all the forms that perturbation theory assumes in the various fields of physics would be vain. We will limit to illustrate its role and its methods in quantum mechanics, which is perhaps the field where it has reached its most mature development and finds its widest applications.

## Introduction

An early example of the use of perturbation theory which clearly illustrates its main ideas is offered by the study of the free fall of a body (Landau and Lifshitz 1960). The equation of motion is

$$\dot{\vec{v}} = \vec{g} + 2\vec{v} \times \vec{\Omega} + \vec{\Omega} \times (\vec{r} \times \vec{\Omega}), \quad (1)$$

where  $\vec{g}$  is the constant gravity acceleration and  $\vec{\Omega}$  the angular velocity of the rotation of the earth about its axis.  $\Omega$  is the parameter characterizing the perturbation. If we wish to find the eastward deviation of the trajectory to first order in  $\Omega$ , we can neglect the third term in the RHS of Eq. 1, whose main effect is to cause a southward deviation (in the northern hemisphere). The ratio of the second term to the first one in the RHS of Eq. 1 (the effective perturbation parameter) is  $\Omega\sqrt{h/g} \simeq 10^{-4}$  for the fall from a height  $h \sim 100$  m, so we can find the effect of  $\Omega$  by writing  $\vec{v} = \vec{v}_0 + \vec{v}_1$  in Eq. 1, where  $\vec{v}_0$  is the zeroth-order solution ( $\vec{v}_0 = \vec{g}t$  if  $\vec{v}_0(0) = 0$ ) and  $\vec{v}_1$  obeys

$$\dot{\vec{v}}_1 = 2\vec{v}_0 \times \vec{\Omega} = 2t\vec{g} \times \vec{\Omega}. \quad (2)$$

The solution is  $\vec{r} = \vec{h} + \frac{1}{2}\vec{g}t^2 + \frac{1}{3}t^3\vec{g} \times \vec{\Omega}$ . The eastward deviation is the deviation in the direction of  $\vec{g} \times \vec{\Omega}$ , and its value is  $\delta = \frac{1}{3}\bar{t}^3 g \Omega \cos \theta$ , where  $\theta$  is the latitude and  $\bar{t}$  the zeroth-order time of fall,  $\bar{t} = \sqrt{2h/g}$ .

While the above example is a nice illustration of the main features of perturbation theory (identification of a perturbation parameter whose powers classify the contributions to the solution, existence of a zeroth-order exact solution), the beginning of modern perturbation theory can be traced back to the work of Rayleigh on the theory of sound (Rayleigh 1894–1896). In essence, he wondered how the normal modes of a vibrating string

$$\rho(x) \frac{\partial^2 v}{\partial t^2} = \frac{\partial^2 v}{\partial x^2} \quad v(0, t) = v(\pi, t) = 0 \quad (3)$$

are modified when passing from a constant density  $\rho = 1$  to a perturbed density  $\rho + \epsilon\sigma(x)$ . To

solve this problem, he wrote down most of the formulae (Rayleigh 1894–1896; Courant and Hilbert 1989) which are still in use to calculate the first-order correction to nondegenerate energy levels in quantum mechanics.

The equation for the normal modes is

$$u''(x) + \lambda\rho(x)u(x) = 0, \quad u(0) = u(\pi) = 0. \quad (4)$$

Let  $u_n^{(0)} \equiv \sqrt{2/\pi} \sin nx$  be the unperturbed solution for the  $n^{\text{th}}$  mode,  $\lambda_n = n^2$ , and  $u_n^{(0)} + \epsilon u_n^{(1)}$  the perturbed solution through first order, corresponding to a frequency  $\lambda_n + \epsilon\mu_n$ . By writing the equation for  $u_n^{(1)}$

$$\frac{d^2 u_n^{(1)}}{dx^2} + \lambda_n u_n^{(1)} + \mu_n u_n^{(0)} + \lambda_n \sigma u_n^{(0)} = 0 \quad (5)$$

$$u_n^{(1)}(0) = u_n^{(1)}(\pi) = 0$$

after multiplying by  $u_r^{(0)}$  and using Green's theorem, he found

$$\mu_n = -\lambda_n \int_0^\pi \sigma(x) \left(u_n^{(0)}\right)^2 dx, \quad (6)$$

$$a_{rn} \equiv \int_0^\pi u_r^{(0)} u_n^{(1)} dx = \frac{\lambda_n}{\lambda_r - \lambda_n} \int_0^\pi \sigma u_r^{(0)} u_n^{(0)} dx \quad (r \neq n) \quad (7)$$

$$\int_0^\pi u_n^{(0)} u_n^{(1)} dx = 0. \quad (8)$$

As an application, Rayleigh found the position  $\pi/2 + \delta x \equiv \pi/2 + \epsilon\tau$  of the nodal point of the perturbed mode  $n = 2$  when the perturbation to the density is  $\sigma = \kappa\delta(x - \pi/4)$ . The vanishing of  $u_2^{(0)} + \epsilon u_2^{(1)}$  determines  $2\sqrt{2/\pi}\tau = u_2^{(1)}(\pi/2)$ . By Eq. 7, the function  $u_2^{(1)}$  has an expansion  $u_2^{(1)} = \sum_{n \neq 2} a_{n2} u_n^{(0)}$ ,  $a_{n2} = \frac{4\kappa}{n^2 - 4} \sin n\pi/4$ . The result for  $\tau$  is

$$\tau = -\frac{2\kappa}{\pi\sqrt{2}} \left(1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \dots\right) = -\frac{\kappa}{2}.$$

(The series in brackets is equal to  $\int_0^1 \frac{1+x^2}{1+x^4} dx = \frac{1}{2} \int_0^\infty \frac{1+x^2}{1+x^4} dx$ , which can be calculated by contour integration.)

Perturbation theory was revived by Schrödinger, who introduced it into quantum mechanics in a pioneering work of 1926 (Schrödinger 1926). There, he applied the concepts and methods which Rayleigh had put forth to the case where the zeroth-order problem was a partial differential equation with nonconstant coefficients, and he wrote down, in the language of wave mechanics, all the relevant formulae which yield the correction to the energy levels and to the wave functions for the case of both nondegenerate and degenerate energy levels. As an application, he calculated the shift of the energy levels of the hydrogen atom in a constant electric field by two different methods. First he observed that in parabolic coordinates, the wave equation is separable also with a constant electric field, which implies that in the subspace of the states with equal zeroth-order energy, the perturbation is diagonal in the basis of the parabolic eigenfunctions, thus circumventing the intricacies of the degenerate case. Later, he used the spherical coordinates, which entail a nondiagonal perturbation matrix and call for the full machinery of the perturbation theory for degenerate eigenvalues.

It is of no use to repeat here Schrödinger's calculations, since the methods which they use are at the core of modern perturbation theory, which is referred to as the Rayleigh-Schrödinger (RS) perturbation theory. It rapidly superseded other approaches (such as that by Born et al. (1926), who worked in the framework of the matrix quantum mechanics) and will be presented in the following sections.

## Presentation of the Problem and an Example

The most frequent application of perturbation theory in quantum mechanics is the approximate calculation of point spectra. The Hamiltonian  $H$  is split into an exactly solvable part  $H_0$  (the unperturbed Hamiltonian) plus a term  $V$  (the perturbation) which, in a sense to be specified later, is small with respect to  $H_0$ :  $H = H_0 + V$ . In many cases, the perturbation contains an adjustable parameter which depends on the actual physical setting. For example, for a

system in an external field, this parameter is the field strength. For weak fields, one expects the spectrum of  $H$  to differ only slightly from the spectrum of  $H_0$ . In these cases, it is convenient to single out the dependence on a parameter by setting

$$H(\lambda) \equiv H_0 + \lambda V. \quad (9)$$

Accordingly, we will write the Schrödinger equation as

$$H(\lambda)\psi(\lambda) = E(\lambda)\psi(\lambda). \quad (10)$$

We will retain the form Eq. 9 of the Hamiltonian even when  $H$  does not contain a variable parameter, thereby understanding that the actual eigenvalues and eigenvectors are the values at  $\lambda = 1$ .

The basic idea of the RS perturbation theory is that the eigenvalues and eigenvectors of  $H$  can be represented as power series

$$\psi(\lambda) = \sum_0^\infty \lambda^n \psi^{(n)} \quad E(\lambda) = \sum_0^\infty \lambda^n \epsilon_n, \quad (11)$$

whose coefficients are determined by substituting expansions Eq. 11 into Eq. 10 and equating terms of equal order in  $\lambda$ . Generally, only the first few terms of the series can be explicitly computed, and the primary task of the RS perturbation theory is their calculation. The practicing scientist who uses perturbation theory never has to tackle the mathematical problem of the convergence of the series. This problem, however, or more generally the connection between the truncated perturbation sums and the actual values of the energy and the wave function is fundamental for the consistency of perturbation theory and will be touched upon in a later section.

Before expounding the technique of the RS perturbation theory, we will consider a simple (two-dimensional) problem which can be solved exactly, since in its discussion several features of perturbation theory will emerge clearly, concerning both the behavior of the energy  $E(\lambda)$  and the behavior of the Taylor expansion of this function. From the physical point of view,



a system with two-dimensional Hilbert space  $\mathbb{C}^2$  can be thought of as a particle with spin 1/2 when the translational degrees of freedom are ignored.

Let us write the Hamiltonian  $H = H_0 + \lambda V$  in a representation where  $H_0$  is diagonal:

$$H = \begin{pmatrix} E_1^0 & 0 \\ 0 & E_2^0 \end{pmatrix} + \lambda \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad (12)$$

$$= \begin{pmatrix} E_1^0 + \lambda V_{11} & \lambda V_{12} \\ \lambda V_{12}^* & E_2^0 + \lambda V_{22} \end{pmatrix}.$$

We consider first the case  $E_1^0 \neq E_2^0$ ,  $V_{12} \neq 0$ . The exact eigenvalues  $E_{1,2}(\lambda)$  of  $H$  are found by solving the secular equation:

$$E_{1,2}(\lambda) = \frac{1}{2} \left[ (E_1^0 + \lambda V_{11}) + (E_2^0 + \lambda V_{22}) \right. \\ \left. \pm \sqrt{\Delta(\lambda)} \right], \quad (13)$$

$$\Delta(\lambda) \equiv ((E_1^0 + \lambda V_{11}) - (E_2^0 + \lambda V_{22}))^2 \\ + 4\lambda^2 |V_{12}|^2. \quad (14)$$

The corresponding eigenvectors, in the so-called intermediate normalization defined by  $(\psi(0), \psi(\lambda)) = 1$ , are

$$\psi_1(\lambda) = \left( 1, \frac{\sqrt{\Delta(\lambda)} - (E_1^0 - E_2^0) - \lambda(V_{11} - V_{22})}{2\lambda V_{12}} \right) \quad (15)$$

$$\psi_2(\lambda) = \left( -\frac{\sqrt{\Delta(\lambda)} - (E_1^0 - E_2^0) - \lambda(V_{11} - V_{22})}{2\lambda V_{21}}, 1 \right). \quad (16)$$

Expanding  $E_{1,2}(\lambda)$  through order  $\lambda^3$ , we get

$$E_1(\lambda) = E_1^0 + \lambda V_{11} + \lambda^2 \frac{|V_{12}|^2}{E_1^0 - E_2^0} \\ - \lambda^3 \frac{|V_{12}|^2 (V_{11} - V_{22})}{(E_1^0 - E_2^0)^2} + O(\lambda^4) \quad (17)$$

$$E_2(\lambda) = E_2^0 + \lambda V_{22} - \lambda^2 \frac{|V_{12}|^2}{E_1^0 - E_2^0} \\ + \lambda^3 \frac{|V_{12}|^2 (V_{11} - V_{22})}{(E_1^0 - E_2^0)^2} \\ + O(\lambda^4). \quad (18)$$

At order 1, only the diagonal matrix elements of  $V$  contribute to  $E_{1,2}$ . The validity of the approximation requires  $\lambda |V_{12}| \ll |E_1^0 - E_2^0|$ . If this condition is not satisfied, that is, if the eigenvalues  $E_1^0, E_2^0$  are “quasi-degenerate,” all terms of the expansion can be numerically of the same order of magnitude and no approximation of finite order makes sense.

Note that within the first-order approximation, “level crossing” ( $E_1(\lambda) = E_2(\lambda)$ ) occurs at

$$\bar{\lambda} = -(E_1^0 - E_2^0)/(V_{11} - V_{22}). \quad (19)$$

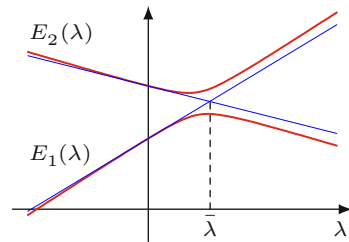
On the other hand, Eq. 13 shows that level crossing is impossible, unless  $V_{12} = 0$ , in which case the first-order approximation yields the exact result. If  $V_{12} \neq 0$ , the behavior of the levels  $E_1(\lambda)$  and  $E_2(\lambda)$  near  $\lambda$  is shown in Fig. 1: the two levels “repel” each other (von Neumann and Wigner 1929).

At first order, the eigenvectors  $\psi_{1,2}(\lambda)$  are

$$\psi_1^{[1]} = (1, -\lambda V_{21}/(E_2^0 - E_1^0)) \quad (20)$$

$$\psi_2^{[1]} = (-\lambda V_{12}/(E_1^0 - E_2^0), 1). \quad (21)$$

The expectation value  $(\psi_1^{[1]}, H\psi_1^{[1]})/(\psi_1^{[1]}, \psi_1^{[1]})$  of the Hamiltonian over  $\psi_1^{[1]}$ , for example, is



**Perturbation Theory in Quantum Mechanics, Fig. 1** The behavior of the exact eigenvalues  $E_{1,2}(\lambda)$  when  $V_{12} = 0$  (blue lines) and when  $V_{12} \neq 0$  (red lines)

$$\begin{aligned}
& E_1^0 + \lambda V_{11} + \lambda^2 \frac{|V_{12}|^2}{E_1^0 - E_2^0} \\
& - \lambda^3 \frac{|V_{12}|^2 (V_{11} - V_{22})}{(E_1^0 - E_2^0)^2} \\
& - \lambda^4 \frac{|V_{12}|^4}{(E_1^0 - E_2^0)^3} + O(\lambda^5) \quad (22)
\end{aligned}$$

which agrees with  $E_1(\lambda)$  up to the  $\lambda^3$  terms (the correct fourth-order term contains also  $|V_{12}|^2(V_{11} - V_{22})^2/(E_1^0 - E_2^0)^3$ ). This is an example of Wigner's  $(2n + 1)$ -theorem (Wigner 1935) (see section "Wigner's Theorem").

The power expansions of  $E_{1,2}(\lambda)$  and  $\psi_{1,2}(\lambda)$  converge in the disk  $|\lambda| < |E_1^0 - E_2^0|/\sqrt{(V_{11} - V_{22})^2 + 4|V_{12}|^2}$ . The denominator is just twice the infimum over  $a$  of the operator norm of  $V - aI$ . Since adding to  $V$  a multiple of the identity does not affect the convergence properties of the Taylor's series of  $E(\lambda)$ , we see that the convergence domain always contains the disk  $|\lambda| < |E_1^0 - E_2^0|/2\|V\|$ , a property which holds true for any bounded perturbation in Hilbert space (see section "Problems with the Perturbation Series").

If  $H_0$  is degenerate, that is,  $E_1^0 = E_2^0 \equiv E^0$ , then the eigenvalues are obtained by diagonalizing  $V$ . The degeneracy is removed, and the corrections to the eigenvalues are of first order in  $\lambda$ :

$$E_{1,2}(\lambda) = E^0 + \frac{1}{2}\lambda \left( V_{11} + V_{22} \pm \sqrt{(V_{11} - V_{22})^2 + 4|V_{12}|^2} \right) \quad (23)$$

while the eigenvectors are  $\lambda$  independent.

The infinite dimensional case is much more involved. In particular, in most cases the perturbation series does not converge at all, that is, its radius of convergence vanishes. However, we shall meet again the three situations discussed above: the case of nondegenerate eigenvalues  $E_n^0$  such that  $|E_n^0 - E_m^0| \gg |\lambda V_{nm}|$ , the case of degenerate eigenvalues, and finally the case of "quasi-degenerate" eigenvalues, i.e., groups of eigenvalues  $E_{n_i}^0$  such that  $|E_{n_i}^0 - E_{n_j}^0| \lesssim |\lambda V_{n_i n_j}|$ . As discussed above, in this last case  $H_{n_i n_j}$  must be diagonalized exactly prior to applying perturbation theory.

## Perturbation of Point Spectra: Nondegenerate Case

In this section, we consider an eigenvector  $\psi_0$  of  $H_0$  belonging to a nondegenerate eigenvalue  $E_0$  and apply the RS theory to determine the power expansions Eq. 11 such that Eq. 10 is satisfied, the Hamiltonian  $H(\lambda)$  being given by Eq. 9. The case of a degenerate eigenvalue will be considered in section "Perturbation of Point Spectra: Degenerate Case." For both cases, the starting point is the substitution of the expansions Eq. 11 into Eq. 10, which, upon equating terms with equal powers, yields the following system of equations:

$$\begin{aligned}
(H_0 - E_0)\psi^{(n)} + V\psi^{(n-1)} &= \sum_{k=0}^{n-1} \psi^{(k)} \epsilon_{n-k}, \quad (24) \\
n &= 1, 2, \dots
\end{aligned}$$

A perturbative calculation of the energy and the wave function through order  $h$  amounts to calculating  $\epsilon_n$  and  $\psi^{(n)}$  up to  $n = h$  and truncating the series in Eq. 11 at  $n = h$ .

## Corrections to the Energy and the Eigenvectors

In the following, let  $\psi_k$ ,  $E_k$  be the normalized eigenvectors and the eigenvalues of  $H_0$ , and let  $\Delta E_{k0} \equiv E_k - E_0$ ,  $V_{hk} \equiv (\psi_h, V\psi_k)$ . The correction  $\epsilon_n$  is recursively defined in terms of the lower-order corrections to the energy and the wave function: by left multiplying Eq. 24 by  $\psi_0$ , we find

$$\epsilon_n = (\psi_0, V\psi^{(n-1)}) - \sum_{h=1}^{n-1} (\psi_0, \psi^{(h)}) \epsilon_{n-h}. \quad (25)$$

Similarly, the components  $(\psi_k, \psi^{(n)})$ ,  $k \neq 0$ , are found by left multiplying by  $\psi_k$ ,  $k \neq 0$ :

$$\begin{aligned}
(\psi_k, \psi^{(n)}) &= -(\psi_k, V\psi^{(n-1)}) \Delta E_{k0}^{-1} \\
&+ \sum_{h=1}^{n-1} (\psi_k, \psi^{(h)}) \epsilon_{n-h} \Delta E_{k0}^{-1}. \quad (26)
\end{aligned}$$

Note that in Eq. 10, the energy  $E(\lambda)$ , hence the corrections  $\epsilon_k$ , is intrinsically defined. On the other hand,  $\psi(\lambda)$  is defined only up to a multiplicative factor  $\alpha(\lambda)$ , which remains undetermined even after normalization is imposed:  $\alpha(\lambda) = \exp(i\varphi(\lambda))$ ,  $\varphi(\lambda) \in \mathbb{R}$ . As a consequence, the wave function corrections  $\psi^{(n)}$  are not uniquely determined by Eq. 24; given the functions  $\psi^{(k)}$  for  $k < n$ , still  $(\psi_0, \psi^{(n)})$  is intrinsically undefined, since to any solution  $\psi^{(n)}$  of Eq. 24, we are allowed to add any multiple of  $\psi_0$ . In terms of the phase factor, this ambiguity stems from the mixing of different  $\psi^{(k)}$  in the expansion of  $\exp(i\varphi(\lambda))\psi(\lambda)$  (Epstein 1954). In particular, one can see that in the process of iteratively inserting Eq. 26 into Eq. 25, the terms multiplying  $(\psi_0, \psi^{(k)})$  must cancel out by noting that under the transformation  $\psi(\lambda) \mapsto \exp(i\lambda^q)\psi(\lambda)$  ( $\psi_0, \psi^{(k)}$ ) is unchanged for  $k < q$ , and  $(\psi_0, \psi^{(q)}) \mapsto (\psi_0, \psi^{(q)}) + i$ . An explicit verification that  $\epsilon_n$  is unaffected by the ambiguities involved in the corrections  $\psi^{(k)}$  was given in Epstein (1968).

Thus, by requiring that in the expression of  $\epsilon_n$ , Eq. 25, the dependence on the values of  $(\psi_0, \psi^{(k)})$ ,  $k \leq n-1$  disappears, we can derive the expression of  $\epsilon_k$  with  $k \leq n$ . For example, after writing

$$\begin{aligned} (\psi_0, V\psi^{(n-1)}) &= V_{00}(\psi_0, \psi^{(n-1)}) \\ &+ \sum_{h \neq 0} V_{0h}(\psi_h, \psi^{(n-1)}), \end{aligned}$$

the independence of  $(\psi_0, \psi^{(n-1)})$  implies  $\epsilon_1 = V_{00}$ . Next, requiring  $\epsilon_n$  to be independent of  $(\psi_0, \psi^{(n-2)})$  determines  $\epsilon_2$  and so on, until finally Eq. 25 gives  $\epsilon_n$ . As an example, we carry through this procedure for  $n = 3$ . Starting from

$$\begin{aligned} \epsilon_3 &= V_{00}(\psi_0, \psi^{(2)}) + \sum_{k \neq 0} V_{0k}(\psi_k, \psi^{(2)}) \\ &- \epsilon_1(\psi_0, \psi^{(2)}) - \epsilon_2(\psi_0, \psi^{(1)}), \end{aligned}$$

we first find

$$\epsilon_1 = V_{00}. \quad (27)$$

Next, from Eq. 26 for  $n = 2$ , we get

$$\begin{aligned} \epsilon_3 &= - \sum_{k \neq 0} \frac{|V_{0k}|^2}{\Delta E_{k0}} (\psi_0, \psi^{(1)}) \\ &- \sum_{h, k \neq 0} \frac{V_{0k}}{\Delta E_{k0}} V_{kh} (\psi_h, \psi^{(1)}) \\ &+ \sum_{k \neq 0} \frac{V_{0k}}{\Delta E_{k0}} (\psi_k, \psi^{(1)}) \epsilon_1 - \epsilon_2 (\psi_0, \psi^{(1)}). \end{aligned}$$

The independence of  $(\psi_0, \psi^{(1)})$  implies

$$\epsilon_2 = - \sum_{k \neq 0} \frac{|V_{0k}|^2}{\Delta E_{k0}}. \quad (28)$$

Finally, by using Eq. 26 for  $n = 1$ , we find

$$\epsilon_3 = \sum_{h, k \neq 0} \frac{V_{0k}}{\Delta E_{k0}} \frac{V_{kh}}{\Delta E_{h0}} V_{h0} - \epsilon_1 \sum_{k \neq 0} \frac{|V_{0k}|^2}{\Delta E_{k0}^2}. \quad (29)$$

If only  $\epsilon_n$  is required, the lower-order corrections being known, one can use a simplified version of Eqs. 25 and 26 where the terms  $(\psi_0, \psi^{(k)})$  are omitted since the beginning. The procedure is easily described if we introduce the vectors  $L$ ,  $R$ , and  $\delta^q$  and the matrix  $D$ :

$$\begin{aligned} L_k &= V_{0k} \quad R_k = V_{k0} \quad \delta_k^q = (\psi_k, \psi^{(q)}) \\ D_{ik} &= \delta_{ik} \Delta E_{k0}^{-1} \quad i, k \geq 1. \end{aligned} \quad (30)$$

Then,  $\epsilon_n = (L, \delta^{n-1})$ , where  $\delta^q$  is given by the recursive relation

$$\begin{aligned} \delta^q &= -DV\delta^{q-1} + \epsilon_1 D\delta^{q-1} + \epsilon_2 D\delta^{q-2} \\ &+ \dots + \epsilon_{q-1} D\delta^1 \end{aligned} \quad (31)$$

with the initial condition  $\delta_k^1 = (\psi_k, \psi^{(1)}) = -\Delta E_{k0}^{-1} V_{k0}$ . Equation 31 can be easily solved with a program of symbolic computation. Also, explicit rules for writing  $\epsilon_n$  directly in terms of matrix elements of the potential can be found in (Bracci and Picasso 2012), where explicit expressions of  $\epsilon_n$  up to  $n = 7$  are also given.

After calculating the energy corrections  $\epsilon_k$ ,  $k \leq n$ , Eq. 26 yields  $(\psi_k, \psi^{(n)})$ . Note however that they are uniquely defined only for  $n = 1$ :

$$\left(\psi_k, \psi^{(1)}\right) = -\frac{V_{k0}}{\Delta E_{k0}}, \quad (32)$$

while the higher-order corrections depend on the choice of  $(\psi_0, \psi^{(k)})$  for lower orders.

For example, by suitably choosing the arbitrary factor  $\alpha(\lambda)$  we referred to after Eq. 26, we can impose  $(\psi_0, \psi(\lambda)) = 1$ . With this choice (known as the “intermediate normalization,” since  $\psi(\lambda)$  is not normalized), we have  $(\psi_0, \psi^{(k)}) = 0$  for any  $k > 0$ . As a result, for the wave function through order  $n$ , we find

$$\psi^{[n]} \equiv \psi_0 + \sum_{k=1}^n \lambda^k \psi^{(k)} \equiv \psi_0 + \delta_n \psi \quad (33)$$

with

$$\left(\psi_0, \psi^{[n]}\right) = 1. \quad (34)$$

In the intermediate normalization, the expression of  $\epsilon_n$

$$\epsilon_n = \left(\psi_0, V\psi^{(n-1)}\right) \quad (35)$$

is just the product  $(L, \delta^{(n-1)})$  discussed above. As for the wave function, the value of  $(\psi_k, \psi^{(n-1)})$  can be read immediately in the expression of  $\epsilon_n$ :  $(\psi_k, \psi^{(n-1)})$  and is obtained from  $\epsilon_n$  by omitting in each term the factor  $V_{0k}$  and the sum over  $k$ . For example, the wave function  $\psi^{[2]} = \psi_0 + \lambda\psi^{(1)} + \lambda^2\psi^{(2)}$  in the intermediate normalization by Eqs. 28 and 29 is

$$\begin{aligned} \psi^{[2]} = & \psi_0 - \lambda \sum_{k=1} \psi_k \frac{V_{k0}}{\Delta E_{k0}} + \lambda^2 \sum_{h,k=1} \psi_k \frac{V_{kh}}{\Delta E_{k0}} \frac{V_{h0}}{\Delta E_{h0}} \\ & - \lambda^2 \epsilon_1 \sum_{k=1} \psi_k \frac{V_{k0}}{\Delta E_{k0}^2}. \end{aligned} \quad (36)$$

In order to calculate expectation values, transition probabilities, and so on, one needs the normalized wave function

$$\psi_N^{[n]} = N^{1/2}(\psi_0 + \delta_n \psi) \quad (37)$$

with  $N^{-1} = 1 + (\delta_n \psi, \delta_n \psi)$ .  $N$  can be chosen real. Note that the wave function  $\psi^{[1]}$  is correctly normalized up to first order.

From the above equations, one sees in which sense the perturbation  $V$  must be small with respect to the unperturbed Hamiltonian  $H_0$ : the separation between the unperturbed energy levels must be large with respect to the matrix elements of the perturbation between those levels, and the total correction  $\delta E$  to  $E_0$  should be small with respect to  $|E_i - E_0|$ ,  $E_i$  standing for any other level of the spectrum of  $H_0$ .

### Wigner's Theorem

From Eq. 24, it follows that

$$H\psi^{[n]} = E^{[n]}\psi^{[n]} + O(\lambda^{n+1}),$$

whence one should infer that if  $E$  is the exact energy,  $E - (\psi_N^{[n]}, H\psi_N^{[n]}) = O(\lambda^{n+1})$ . Wigner's result that the knowledge of  $\psi^{[n]}$  allows the calculation of the energy up to order  $2n+1$  (Wigner's  $2n+1$  theorem) is therefore noteworthy (Wigner 1935). Indeed, he proved that if  $E$  is the exact energy,

$$E - \frac{(\psi^{[n]}, H\psi^{[n]})}{(\psi^{[n]}, \psi^{[n]})} = O(\lambda^{2n+2}).$$

To this purpose, let

$$\chi^{(n+1)} = \psi - \frac{\psi^{[n]}}{\sqrt{(\psi^{[n]}, \psi^{[n]})}},$$

where  $\psi$  is the normalized exact wave function,  $H\psi = E\psi$ . Then,

$$\begin{aligned} \chi^{(n+1)} &= O(\lambda^{n+1}), \quad (\psi, \chi^{(n+1)}) + (\chi^{(n+1)}, \psi) \\ &= -(\chi^{(n+1)}, \chi^{(n+1)}) = O(\lambda^{2n+2}). \end{aligned}$$

As a consequence

$$\frac{(\psi^{[n]}, H\psi^{[n]})}{(\psi^{[n]}, \psi^{[n]})} - (\psi, H\psi) = O(\lambda^{2n+2}).$$

We make explicit this point with an example. Since

$$\psi_N^{[1]} = \frac{\psi_0 + \lambda\psi^{(1)}}{\sqrt{1 + \lambda^2(\psi^{(1)}, \psi^{(1)})}},$$

by using Eq. 32 and recalling Eqs. 28 and 29, we have

$$\begin{aligned} (\psi_N^{[1]}, H\psi_N^{[1]}) &= E_0 + \lambda\epsilon_1 + \frac{\lambda^2\epsilon_2 + \lambda^3\epsilon_3}{1 + \lambda^2(\psi^{(1)}, \psi^{(1)})} \\ &= E_0 + \lambda\epsilon_1 + \lambda^2\epsilon_2 + \lambda^3\epsilon_3 + O(\lambda^4). \end{aligned}$$

### The Feynman-Hellmann Theorem

The RS perturbative expansion rests on the hypothesis that both the eigenvalues  $E(\lambda)$  and the corresponding eigenvectors  $\psi(\lambda)$  admit a power series expansion, in short, that they are analytic functions of  $\lambda$  in a neighborhood of the origin. As we shall see in section “Problems with the Perturbation Series,” as a rule it is not so, and the perturbative expansion gives rise only to a formal series. For this reason, it is advisable to derive the various terms of the perturbation expansion without assuming analyticity. If we need  $E(\lambda)$  and  $\psi(\lambda)$  through order  $n$ , it is sufficient to assume that, as functions of  $\lambda$ , they are  $C^{n+1}$ , that is, continuously differentiable  $(n+1)$  times. The procedure consists in taking the derivatives of Eq. 10 (Epstein 1954; Krieger 1968): at the first step, we get

$$H\psi'(\lambda) + V\psi(\lambda) = E'(\lambda)\psi(\lambda) + E(\lambda)\psi'(\lambda) \quad (38)$$

and by left multiplication by  $\psi(\lambda)$ , with  $(\psi(\lambda), \psi(\lambda)) = 1$ , we get

$$E'(\lambda) = (\psi(\lambda), V\psi(\lambda)), \quad (39)$$

which is a special case of the Feynman-Hellmann theorem (Hellmann 1937; Feynman 1939):

$$\frac{\partial E}{\partial \lambda} = \left( \psi(\lambda), \frac{\partial H}{\partial \lambda} \psi(\lambda) \right). \quad (40)$$

For  $\lambda = 0$ , we find

$$E'(0) = V_{00}, \quad (41)$$

whence  $\epsilon_1 = V_{00}$ , in agreement with Eq. 27. Next, after left multiplying Eq. 38 by  $\psi_k$  and taking  $\lambda = 0$ , we get

$$(\psi_k, \psi') = -\frac{V_{k0}}{\Delta E_{k0}} \quad (42)$$

which, again, agrees with Eq. 32. Taking now the derivative of Eq. 39 at  $\lambda = 0$  and using Eq. 41, we obtain

$$E''(0) = 2 \sum_{k=1} V_{0k}(\psi_k, \psi') = -2 \sum_{k=1} \frac{|V_{0k}|^2}{\Delta E_{k0}} \quad (43)$$

whence  $\epsilon_2 = \frac{1}{2}E''(0)$ , in agreement with Eq. 28.

It is clear that the procedure can be pursued to any allowed order and that the results for the energy corrections, as well as for the wave functions, are the same as those we obtained earlier by the RS technique. However, the conceptual difference that no analyticity hypothesis is required is important since in many cases this hypothesis is not satisfied.

As to the relation of  $E^{[n]} \equiv E_0 + \lambda\epsilon_1 + \dots + \lambda^n\epsilon_n$  with  $E(\lambda)$ , we recall that, since by assumption  $E(\lambda)$  is  $C^{n+1}$ , we can write Taylor's formula with a remainder:

$$E(\lambda) = \sum_0^n \frac{E^{(p)}}{p!} \lambda^p + \frac{E^{(n+1)}(\theta\lambda)}{(n+1)!} \lambda^{n+1}, \quad 0 < \theta < 1. \quad (44)$$

As observed in Krieger (1968), since for small  $\lambda$  the sign of the remainder is the sign of  $E^{(n+1)}(0)\lambda^{n+1}$ , Eq. 44 allows to establish whether the sum in Eq. 44 underestimates or overestimates  $E(\lambda)$ . Moreover, if two consecutive terms, say  $q$  and  $q+1$ , have opposite sign, then (for

sufficiently small  $\lambda$ )  $E(\lambda)$  is bracketed between the partial sums including and excluding the  $q$ th term. It is a pity that no one can anticipate how small such a  $\lambda$  should be. (Of course these remarks apply to the RS truncated series as well.)

### Perturbation of Point Spectra: Degenerate Case

The case when the unperturbed energy  $E_0$  is a degenerate eigenvalue of  $H_0$ , i.e., in the Hilbert space there exists a subspace  $W_0$  generated by a set  $\{\psi_0^{(i)}\}$ ,  $1 \leq i \leq n_0$ , of orthogonal normalized states, such that each  $\psi_0$  in  $W_0$  obeys  $(H_0 - E_0)\psi_0 = 0$ , deserves a separate treatment. The main problem is that, if  $\psi(\lambda)$  is an eigenstate of the exact Hamiltonian  $H = H_0 + \lambda V$ , we do not know beforehand which state of  $W_0$   $\psi(0)$  is.

In order to use a more compact notation, it is convenient to introduce the projection  $P_0$  onto the subspace  $W_0$  and its complement  $Q_0$ :

$$P_0\psi = \sum_{i=0}^{n_0} \psi_0^{(i)} \left( \psi_0^{(i)}, \psi \right) \quad Q_0 \equiv I - P_0, \quad (45)$$

where  $\psi_0^{(i)}$ ,  $1 \leq i \leq n_0$ , is any orthonormal basis of  $W_0$ . The Hamiltonian  $H = H_0 + \lambda V$  can be written as

$$\begin{aligned} H &= (P_0 + Q_0)(H_0 + \lambda V)(P_0 + Q_0) \\ &= E_0 P_0 + \lambda V_{PP} + \lambda V_{PQ} + \lambda V_{QP} \\ &\quad + \lambda V_{QQ} + Q_0 H_0 Q_0, \end{aligned} \quad (46)$$

where

$$\begin{aligned} V_{PP} &= P_0 V P_0, V_{PQ} = P_0 V Q_0, \\ V_{QP} &= Q_0 V P_0, V_{QQ} = Q_0 V Q_0. \end{aligned} \quad (47)$$

After projecting the Schrödinger equation onto  $W_0$  and its orthogonal complement  $W_0^\perp$ , we find

$$(E_0 + \lambda V_{PP})P_0\psi + \lambda V_{PQ}Q_0\psi = EP_0\psi \quad (48)$$

$$\lambda V_{QP}P_0\psi + Q_0 H_0 Q_0\psi + \lambda V_{QQ}Q_0\psi = EQ_0\psi. \quad (49)$$

Letting

$$H_{QQ} = Q_0 H Q_0 = Q_0 H_0 Q_0 + \lambda V_{QQ}, \quad (50)$$

$Q_0\psi$  can be extracted from Eq. 49:

$$Q_0\psi = \lambda(E - H_{QQ})^{-1} V_{QP} P_0\psi. \quad (51)$$

Note that in Eq. 49, the operator  $H_{QQ}$  acts on vectors of  $W_0^\perp$  and that  $E - H_{QQ}$  does possess an inverse in  $W_0^\perp$ . Indeed, the existence of a vector  $\zeta$  in  $W_0^\perp$  such that

$$(H_{QQ} - E)\zeta = 0 \quad (52)$$

contradicts the assumptions which perturbation theory is grounded in: the separation between  $E(\lambda)$  and  $E(0)$  should be negligible with respect to the separation between different eigenvalues of  $H_0$ . Actually, if  $\psi_k$  is such that  $H_0\psi_k = E_k\psi_k$ ,  $E_k \neq E_0$ , by left multiplying Eq. 52 by  $\psi_k$ , we would find

$$(E - E_k)(\psi_k, \zeta) = \lambda(\psi_k, V\zeta),$$

where the LHS is of order 0 in  $\lambda$ , whereas the RHS of order 1.

By substituting Eq. 51 into Eq. 48, we have

$$\begin{aligned} (E_0 + \lambda V_{PP})P_0\psi + \lambda^2 V_{PQ}(E - H_{QQ})^{-1} V_{QP} P_0\psi \\ = EP_0\psi. \end{aligned} \quad (53)$$

The energy shifts  $\Delta E \equiv E - E_0$  appear as eigenvalues of an operator  $A(E)$  acting in  $W_0$ :

$$A(E) \equiv \lambda V_{PP} + \lambda^2 V_{PQ}(E - H_{QQ})^{-1} V_{QP} \quad (54)$$

which however still depends on the unknown exact energy  $E$ . A calculation of the energy corrections up to a given order is possible, starting from Eq. 54, provided we expand the term  $(E - H_{QQ})^{-1}$  as far as is necessary to include all terms of the requested order.

### Corrections to the Energy and the Eigenvectors

The contributions  $\epsilon_i$  are extracted from Eq. 53 by expanding

$$\begin{aligned} E &= E_0 + \lambda\epsilon_1 + \lambda^2\epsilon_2 + \dots \\ P_0\psi &= \varphi_0 + \lambda\varphi_1 + \lambda^2\varphi_2 + \dots \end{aligned}$$

and equating terms of equal order. At the first order, since the second term in the LHS of Eq. 53 is of order 2 or larger, we have

$$V_{PP}\varphi_0 = \epsilon_1\varphi_0. \quad (55)$$

The first-order corrections to the energy are the eigenvalues of the matrix  $V_{PP}$ , and the corresponding zeroth-order wave function is the corresponding eigenvector.

In the most favorable case, the eigenvalues of  $V_{PP}$  are simple, and the degeneracy is completely removed since the first order of perturbation theory. In this case, in order to get the higher-order corrections, we can avail ourselves of the arbitrariness in the way of splitting the exact Hamiltonian into a solvable unperturbed Hamiltonian plus a perturbation by putting

$$\begin{aligned} H &= (H_0 + \lambda V_{PP}) + \lambda(V - V_{PP}) \\ &\equiv H'_0 + \lambda V'. \end{aligned} \quad (56)$$

The eigenvectors of  $H'_0$  are the solutions of Eq. 55, with eigenvalues  $E_0 + \lambda\epsilon_1^{(i)}$ ,  $1 \leq i \leq n_0$  plus the eigenvectors  $\psi_j$  of  $H_0$  with eigenvalues  $E_j \neq E_0$ . Since the eigenvalues  $E_0 + \lambda\epsilon_1^{(i)}$  are no longer degenerate, the formalism of non-degenerate perturbation theory can be applied, but a warning is in order. When in higher perturbation orders a denominator  $\Delta E_{k0}$  occurs with the index  $k$  referring to another vector of the basis of  $W_0$ , this denominator is of order  $\lambda$ , and consequently the order of the term containing this denominator is lower than the naive  $V$ -counting would imply. In each such term, the effective order is the  $V$ -counting order minus the number of these denominators. As shown below, this situation occurs starting from terms of order 4 in the perturbation  $V$ . Note that, also in the case of

noncomplete removal of the degeneracy, the procedure outlined above, with obvious modifications, can be applied to search the higher-order corrections to those eigenvalues which at first order turn out to be nondegenerate.

If a residual degeneracy still exists, i.e., an eigenvalue  $\epsilon_1$  of Eq. 55 is not simple, we must explore the higher-order corrections until the degeneracy, if possible, is removed. First of all, we must disentangle the contributions of different orders in  $\lambda$  from  $(E - H_{QQ})^{-1}$ . Since

$$\begin{aligned} E - H_{QQ} &= (E - Q_0 H_0 Q_0) \\ &[1 - \lambda(E - Q_0 H_0 Q_0)^{-1} V_{QQ}], \end{aligned}$$

we have

$$\begin{aligned} (E - H_{QQ})^{-1} &= \sum_0^\infty \lambda^n [(E - Q_0 H_0 Q_0)^{-1} V_{QQ}]^n \\ &(E - Q_0 H_0 Q_0)^{-1}. \end{aligned} \quad (57)$$

As the energy  $E$  still contains contributions of any order, the operator  $(E - Q_0 H_0 Q_0)^{-1}$  must in turn be expanded into a series in  $\lambda$ . To make notations more readable, we define

$$\frac{Q_0}{a^n} \equiv (E_0 - Q_0 H_0 Q_0)^{-n}. \quad (58)$$

The second-order terms from Eqs. 53 and 57 give

$$V_{PP}\varphi_1 + V_{PQ} \frac{Q_0}{a} V_{QP}\varphi_0 = \epsilon_2\varphi_0 + \epsilon_1\varphi_1. \quad (59)$$

Let  $P_0^{(i)}$  be the projections onto the subspaces  $W_0^{(i)}$  of  $W_0$  corresponding to the eigenvalues  $\epsilon_1^{(i)}$ :

$$\begin{aligned} P_0 &= \sum_i P_0^{(i)}, V_{PP} = \lambda \sum_i \epsilon_1^{(i)} P_0^{(i)}, P_1 \equiv P_0^{(1)}, \\ \epsilon_1 &\equiv \epsilon_1^{(1)}. \end{aligned} \quad (60)$$

By projecting onto  $W_1 \equiv W_0^{(1)}$  and recalling that  $\varphi_0$  is in  $W_1$ , we get



$$P_1 V_{PQ} \frac{Q_0}{a} V_{QP} \varphi_0 = \epsilon_2 \varphi_0, \quad (61)$$

whence  $\epsilon_2$  is an eigenvalue of the operator

$$V_1 \equiv P_1 V_{PQ} \frac{Q_0}{a} V_{QP} P_1 = P_1 V \frac{Q_0}{a} V P_1. \quad (62)$$

Again, if the eigenvalue  $\epsilon_2$  is nondegenerate, we can use the previous theory by splitting the Hamiltonian as

$$\begin{aligned} H &= (H_0 + \lambda V_{PP} + \lambda V_1) + \lambda(V - V_{PP} - V_1) \\ &\equiv H_0'' + \lambda V''. \end{aligned} \quad (63)$$

The vectors which make  $V_1$  diagonal belong to nondegenerate eigenvalues of  $H_0''$ ; hence, the nondegenerate theory can be applied. If, on the contrary, the eigenvalue  $\epsilon_2$  of  $V_1$  is still degenerate, the above procedure can be carried out one step further, with the aim of removing the residual degeneracy. We work out the calculation for  $\epsilon_3$ , since a new aspect of degenerate perturbation theory emerges: a truly third-order term which is the ratio of a term of order 4 in the potential and a term of first order (see Eq. 69 below).

From Eqs. 53 and 57, we extract the contribution of order 3:

$$\begin{aligned} V_{PP} \varphi_2 + V_{PQ} \frac{Q_0}{a} V_{QP} \varphi_1 - \epsilon_1 V_{PQ} \frac{Q_0}{a^2} V_{QP} \varphi_0 + \\ V_{PQ} \frac{Q_0}{a} V_{QQ} \frac{Q_0}{a} V_{QP} \varphi_0 = \epsilon_1 \varphi_2 + \epsilon_2 \varphi_1 + \epsilon_3 \varphi_0. \end{aligned} \quad (64)$$

We want to convert this equation into an eigenvalue problem for  $\epsilon_3$ . In analogy with Eq. 60, we have

$$\begin{aligned} P_1 = \sum_i P_1^{(i)}, V_1 = \sum \epsilon_2^{(i)} P_1^{(i)}, P_2 \equiv P_1^{(1)}, \\ \epsilon_2 \equiv \epsilon_2^{(1)}. \end{aligned} \quad (65)$$

Since  $P_2 V_{PP} = \epsilon_1 P_2$ , first we eliminate  $\varphi_2$  by applying  $P_2$  to Eq. 64:

$$\begin{aligned} P_2 V_{PQ} \frac{Q_0}{a} V_{QP} \varphi_1 - \epsilon_1 P_2 V_{PQ} \frac{Q_0}{a^2} V_{QP} \varphi_0 + \\ P_2 V_{PQ} \frac{Q_0}{a} V_{QQ} \frac{Q_0}{a} V_{QP} \varphi_0 = \epsilon_3 P_2 \varphi_0 + \epsilon_2 P_2 \varphi_1. \end{aligned} \quad (66)$$

Writing  $\varphi_1 = \sum_i P_0^{(i)} \varphi_1$ , since  $P_2 V_1 = \epsilon_2 P_2$  the contribution with  $i = 1$  of the first term in the LHS of Eq. 66 is  $P_2 V_1 \varphi_1 = \epsilon_2 P_2 \varphi_1$ . Hence, Eq. 66 reads

$$\begin{aligned} \sum_{i \neq 1} P_2 V_{PQ} \frac{Q_0}{a} V_{QP} P_0^{(i)} \varphi_1 - \epsilon_1 P_2 V_{PQ} \frac{Q_0}{a^2} V_{QP} \varphi_0 + \\ P_2 V_{PQ} \frac{Q_0}{a} V_{QQ} \frac{Q_0}{a} V_{QP} \varphi_0 = \epsilon_3 P_2 \varphi_0 + \epsilon_2 \sum_{i \neq 1} P_0^{(i)} \varphi_1. \end{aligned} \quad (67)$$

Finally,  $P_0^{(i)} \varphi_1$ ,  $i \neq 1$ , is extracted from Eq. 59 by projecting with  $P_0^{(i)}$ ,  $i \neq 1$ , and recalling that  $P_0^{(i)} \varphi_0 = 0$  if  $i \neq 1$ :

$$P_0^{(i)} \varphi_1 = P_0^{(i)} V_{PQ} \frac{Q_0}{a} V_{QP} \varphi_0 / (\epsilon_1 - \epsilon_1^{(i)}), \quad i \neq 1. \quad (68)$$

Substituting into Eq. 67, we see that  $\epsilon_3$  is defined by the eigenvalue equation for the operator:

$$\begin{aligned} V_2 \equiv P_2 V \frac{Q_0}{a} V \frac{Q_0}{a} V P_2 - \epsilon_1 P_2 V \frac{Q_0}{a^2} V P_2 + \\ \sum_{i \neq 1} P_2 V \frac{Q_0}{a} V \frac{P_0^{(i)}}{\epsilon_1 - \epsilon_1^{(i)}} V \frac{Q_0}{a} V P_2. \end{aligned} \quad (69)$$

Despite the presence of four factors in the potential, the last term is actually a third-order term due to the denominators  $\epsilon_1 - \epsilon_1^{(i)}$ .

The procedure outlined above, which essentially embodies the Rayleigh-Schrödinger approach, can be pursued until the degeneracy is (if possible, see below section “Symmetry and Degeneracy”) completely removed, after which the theory for the nondegenerate case can be used. Rather than detailing the calculations, we present an alternative iterative procedure due to Bloch (1958) which allows a more systematic



calculation of the corrections to the energy and the wave function.

### Bloch's Method

In Eqs. 53 and 54, we have seen that the energy corrections  $\Delta E$  and the projections onto  $W_0$  of the vectors  $\psi_k(\lambda)$  are eigenvalues and eigenvectors of an operator acting in  $W_0$ . This observation is not immediately useful since the operator depends on the unknown exact energy  $E(\lambda)$ . However, it is possible to produce an operator  $B(\lambda)$ , which can be calculated in terms of known quantities and has the property that, if  $E_k(\lambda)$ ,  $\psi_k(\lambda)$  are eigenvalues and eigenvectors of Eq. 10 such that  $E_k(0) = E_0$ , then

$$B(\lambda)P_0\psi_k(\lambda) = \Delta E_k P_0\psi_k(\lambda). \quad (70)$$

First of all, note that the vectors  $P_0\psi_k(\lambda)$  are a basis for the subspace  $W_0$ . Indeed, it is implicit in the assumption that perturbation theory does work and that the perturbing potential should produce only slight modifications of the unperturbed eigenvectors of the Hamiltonian, so that the vectors  $P_0\psi_k(\lambda)$  are linearly independent (although not orthogonal). Since their number equals the dimension of  $W_0$ , they are a basis for this subspace.

Following Bloch (1958), we define a  $\lambda$ -dependent operator  $U$  in this way:

$$UP_0\psi_k(\lambda) = \psi_k(\lambda); \quad UQ_0 = 0. \quad (71)$$

As a consequence, we have

$$U = UP_0, \quad P_0U = P_0, \quad (72)$$

$$U\psi_k(\lambda) = \psi_k(\lambda). \quad (73)$$

The former of Eq. 72 follows immediately from the definition of  $U$ . Hence  $P_0U = P_0UP_0$ , which implies the latter of Eq. 72. Equation 73 is verified by applying the former of Eqs. 72 to  $\psi_k(\lambda)$ .

Let

$$B(\lambda) \equiv \lambda P_0 V U. \quad (74)$$

We verify that if  $\Delta E_k \equiv E_k - E_0$ , then

$$B(\lambda)P_0\psi_k(\lambda) = \Delta E_k P_0\psi_k(\lambda). \quad (75)$$

Indeed, by Eq. 71 we have  $P_0 V U P_0 \psi_k(\lambda) = P_0 V \psi_k(\lambda)$ . Writing Eq. 10 as

$$(H_0 - E_0 + \lambda V)\psi_k(\lambda) = \Delta E_k \psi_k(\lambda)$$

and multiplying by  $P_0$ , we find

$$\lambda P_0 V \psi_k(\lambda) = \Delta E_k P_0 \psi_k(\lambda); \quad (76)$$

hence, Eq. 75 is satisfied.

A practical use of Eq. 75 requires an iterative definition of  $U$  in terms of known quantities. From Eqs. 71 and 72, we have

$$U = P_0U + Q_0U = P_0 + Q_0UP_0. \quad (77)$$

We calculate the latter term of Eq. 77 on the vectors  $P_0\psi_k(\lambda)$ . Since

$$(\lambda V - \Delta E_k)\psi_k = (E_0 - H_0)\psi_k,$$

recalling Eq. 71 we have

$$\begin{aligned} Q_0UP_0\psi_k(\lambda) &= Q_0\psi_k(\lambda) \\ &= \frac{Q_0}{a}(\lambda V - \Delta E_k)\psi_k(\lambda) \\ &= \lambda \frac{Q_0}{a} V U \psi_k(\lambda) - \Delta E_k \frac{Q_0}{a} U \psi_k(\lambda) \\ &= \lambda \frac{Q_0}{a} V U \psi_k(\lambda) - \Delta E_k \frac{Q_0}{a} U P_0 \psi_k(\lambda). \end{aligned}$$

By Eq. 76,

$$\begin{aligned} Q_0UP_0\psi_k(\lambda) &= \lambda \frac{Q_0}{a} V U \psi_k(\lambda) - \lambda \frac{Q_0}{a} U P_0 V \psi_k(\lambda) \\ &= \lambda \frac{Q_0}{a} V U \psi_k(\lambda) - \lambda \frac{Q_0}{a} U P_0 V U \psi_k(\lambda) \\ &= \lambda \frac{Q_0}{a} (V U - U V U) P_0 \psi_k(\lambda). \end{aligned}$$

As a consequence, the desired iterative equation for  $U$  is

$$U = P_0 + \lambda \frac{Q_0}{a} (V U - U V U). \quad (78)$$

Equation 78 in turn allows an iterative definition of the operator  $B(\lambda)$  of Eq. 74 depending only

on quantities which can be computed in terms of the known spectral representation of  $H_0$ . Knowing  $U$  through order  $n - 1$  gives  $B^{[n]}(\lambda) \equiv \sum_{i=1}^n B^{(i)}(\lambda)$ , whose eigenvalues are the energy corrections through order  $n$ . In fact, if  $B = \sum_{i=1}^{\infty} B^{(i)}$  and  $P_0 \psi_k = \sum_{s=0}^{\infty} \lambda^s \varphi_s$ , the order  $r$  contribution to Eq. 75 is

$$\sum_{i=1}^r B^{(i)} \varphi_{r-i} = \sum_{i=1}^r \epsilon_i \varphi_{r-i}. \quad (79)$$

Defining  $P_0 \psi_k^{[n]} \equiv \sum_0^n \lambda^r \varphi_r \equiv \varphi^{[n]}$ ,  $\Delta E^{[n]} \equiv \sum_1^n \lambda^r \epsilon_r$ , we see that the sum of Eq. 79 for values of  $r$  through  $n$  gives

$$B^{[n]} \varphi^{[n]} = \Delta E^{[n]} \varphi^{[n]} + O(\lambda^{n+1}). \quad (80)$$

Once  $P_0 \psi_k^{[n]}(\lambda)$  has been found, Eq. 71 gives the component of  $\psi_k(\lambda)$  in  $W_0^\perp$  through order  $n + 1$ . As an example, for  $n = 3$  we have

$$\begin{aligned} U^{[2]} &= P_0 + \lambda \frac{Q_0}{a} V P_0 + \lambda^2 \frac{Q_0}{a} V \frac{Q_0}{a} V P_0 \\ &\quad - \lambda^2 \frac{Q_0}{a} V P_0 V P_0, \end{aligned} \quad (81)$$

$$\begin{aligned} B^{[3]} &= \lambda P_0 V P_0 + \lambda^2 P_0 V \frac{Q_0}{a} V P_0 + \\ &\quad \lambda^3 P_0 V \frac{Q_0}{a} V \frac{Q_0}{a} V P_0 - \lambda^3 P_0 V \frac{Q_0}{a^2} V P_0 V P_0. \end{aligned} \quad (82)$$

If  $W_0$  is one-dimensional, Eq. 82 gives for  $\lambda \epsilon_1 + \lambda^2 \epsilon_2 + \lambda^3 \epsilon_3$  the same result as Eqs. 27–29.

The main difference between the RS perturbation theory and Bloch's method is that within the former, the energy corrections through order  $n$  are calculated by means of a sequential computation starting from  $\epsilon_1$ , with the consequence that at each step the dimension of the matrix to be diagonalized is smaller. Conversely, within Bloch's method, one has to diagonalize the matrix  $B^{[n]}(\lambda)$ , which has the dimension of  $W_0$ . However, as noted above, for  $n > 1$  the eigenvalues of  $B^{[n]}(\lambda)$  are different from  $\lambda \epsilon_1 + \lambda^2 \epsilon_2 + \dots + \lambda^n \epsilon_n$  by terms of order at least  $n + 1$ . Similarly, the eigenvectors of Eq. 80 differ from the component

in  $W_0$  of  $\psi^{[n]} = \psi_0 + \lambda \psi^{(1)} + \dots + \lambda^n \psi^{(n)}$  by terms of order larger than  $n$ .

It is instructive to reconsider the calculation of  $\epsilon_2$  and  $\epsilon_3$  in the light of Bloch's method. If  $P_0 = P_1 + P_1'$ , then  $P_0 V P_0 = \epsilon_1 P_1 + P_1' V P_1'$  and

$$\begin{aligned} B^{[2]}(\lambda) &= \lambda \epsilon_1 P_1 + \lambda P_1' V P_1' + \lambda^2 P_1 V \frac{Q_0}{a} V P_1 + \\ &\quad \lambda^2 P_1' V \frac{Q_0}{a} V P_1' + \lambda^2 P_1 V \frac{Q_0}{a} V P_1' + \lambda^2 P_1' V \frac{Q_0}{a} V P_1. \end{aligned}$$

The last two terms represent off-diagonal blocks which can be omitted for the calculation of  $\lambda \epsilon_1 + \lambda^2 \epsilon_2$ , since the lowest-order contribution to the eigenvalues of a matrix  $X$  from the off-diagonal terms  $X_{ij}$  is  $|X_{ij}|^2 / (X_{ii} - X_{jj})$ . For a second-order expansion as  $B^{[2]}$ , this yields third-order contributions of the type

$$\lambda^3 P_1 V \frac{Q_0}{a} V \frac{P_1'}{\epsilon_1 - \epsilon_1'} V \frac{Q_0}{a} V P_1.$$

These are just the contributions to  $\epsilon_3$  which we met in the RS approach: the expression of  $V_2$  given in Eq. 69 combines the block-diagonal term of order 3 with the off-diagonal terms of order 2 giving a third-order contribution.

### The Quasi-degenerate Case

There are cases, in both atomic and molecular physics, where the energy levels of  $H_0$  present a multiplet structure: the energy levels are grouped into “multiplets” whose separation  $\Delta E$  is large compared to the energy separation  $\delta E$  between the levels belonging to the same multiplet. For instance, in atomic physics this is the case of the fine structure (due to the so-called spin-orbit interaction) or of the hyperfine structure (due to the interaction of the nuclear magnetic moment with the electrons); in molecular physics typically, this is the case of the rotational levels associated with the different and widely separated vibrational levels.

If a perturbation  $V$  is such that its matrix elements between levels of the same multiplet are comparable to  $\delta E$ , while being small with respect to  $\Delta E$ , then naive perturbation theory fails because of the small energy denominators pertaining to levels belonging to the same multiplet. To solve

this problem, named the problem of quasi-degenerate levels, once again we can exploit the arbitrariness in the way of splitting the Hamiltonian  $H$  into an unperturbed Hamiltonian and a perturbation. Let

$$\begin{aligned} E_0^{(1)} &\equiv E_0 + \delta E^{(1)}, \\ E_0^{(2)} &\equiv E_0 + \delta E^{(2)}, \dots, \\ E_0^{(n)} &\equiv E_0 + \delta E^{(n)} \end{aligned}$$

be the unperturbed energies within a multiplet, with  $E_0$  any value close to the  $E_0^{(i)}$ 's (for instance, their mean value), and  $P_0^{(i)}$  the projections onto the corresponding eigenspaces. Let

$$H_0^0 \equiv H_0 - \sum_i \delta E^{(i)} P_0^{(i)}, \quad \tilde{V} \equiv \lambda V + \sum_i \delta E^{(i)} P_0^{(i)},$$

so that

$$H = H_0^0 + \tilde{V}. \quad (83)$$

We consider  $H_0^0$  as the unperturbed Hamiltonian and  $\tilde{V}$  as the perturbation. From the physical point of view, this procedure, if applied to all multiplets, is just the inclusion into the perturbation of those terms of  $H_0$  that are responsible for the multiplet structure. With the splitting of the Hamiltonian as in Eq. 83, we can apply the methods of degenerate perturbation theory. The most efficient of these techniques is Bloch's method, which yields a simple prescription for the calculation of the corrections of any order. If, for example, we are content with the lowest order, we must diagonalize the matrix  $P_0 \tilde{V} P_0$  or equivalently  $P_0 H P_0$ , that is, the energies through first order are the eigenvalues of the equation

$$P_0 H P_0 \psi = E P_0 \psi, \quad (84)$$

where  $P_0 = \sum_i P_0^{(i)}$  is the projection onto  $W_0$ , the eigenspace of  $H_0^0$  corresponding to the eigenvalue  $E_0$ . These eigenvalues are algebraic functions of  $\lambda$ , and no finite order approximation is meaningful, since all terms can be numerically of the same order, due to the occurrence of small denominators  $(\delta E^{(i)} - \delta E^{(j)})^n$ .

## The Brillouin-Wigner Method

Equations 54 and 57 yield an alternative approach to the calculation of the energy shift  $\Delta E$  due to a perturbation to a nondegenerate energy level  $E_0$ , the so-called Brillouin-Wigner method (Brillouin 1932; Wigner 1935; Hannabuss 1997). In this case  $W_0$ , the space spanned by the unperturbed eigenvector  $\psi_0$ , is one-dimensional. The correction  $\Delta E$  obeys the equation

$$\Delta E = (\psi_0, A(E) \psi_0), \quad (85)$$

where the operator  $A(E)$  is defined in Eq. 54.

Substituting into the expression of  $A$  the expansion given in Eq. 57 for  $(E - H_{QQ})^{-1}$  and noting that if  $\{E_k\}$  is the spectrum of  $H_0$ ,

$$\begin{aligned} (\psi_0, V_{PQ}(E - Q_0 H_0 Q_0)^{-1} V_{QP} \psi_0) &= \sum_{k \neq 0} \frac{|V_{0k}|^2}{E - E_k}, \\ (\psi_0, V_{PQ}(E - Q_0 H_0 Q_0)^{-1} V_{QQ}(E - Q_0 H_0 Q_0)^{-1} V_{QP} \psi_0) &= \sum_{k, h \neq 0} V_{0k}(E - E_k)^{-1} V_{kh}(E - E_h)^{-1} V_{h0} \end{aligned}$$

and so on, we find the following implicit expression for the exact energy  $E$ :

$$\begin{aligned} E &= E_0 + \lambda(\psi_0, V \psi_0) + \lambda^2 \sum_{k \neq 0} \frac{|V_{0k}|^2}{E - E_k} \\ &+ \lambda^3 \sum_{k, h \neq 0} \frac{V_{0k}}{E - E_k} \frac{V_{kh}}{E - E_h} V_{h0} + \dots \end{aligned} \quad (86)$$

Consistently with the assumption that perturbation theory does work, the denominators in Eq. 86 are nonvanishing. The equation can be solved by arresting the expansion to a given power  $n$  in the potential and searching a solution iteratively starting with  $E = E_0$ . However, the result differs from the energy  $E^{[n]} = E_0 + \lambda \epsilon_1 + \lambda^2 \epsilon_2 + \dots + \lambda^n \epsilon_n$ , calculated by means of the RS perturbation theory, by terms of order  $n + 1$  in the potential. The result of the RS perturbation theory can be recovered from the Brillouin-Wigner approach by substituting in the denominators  $E = E_0 + \lambda \epsilon_1 + \lambda^2 \epsilon_2 + \dots + \lambda^n \epsilon_n$ , expanding the denominators in powers of  $\lambda^k \epsilon_k / E_0$  and

equating terms of equal orders in both sides of Eq. 86.

As for the perturbed wave function, if the intermediate normalization is used, by Eq. 51 we have:

$$\psi = \psi_0 + Q_0\psi = \psi_0 + \lambda(E - H_{Q_0})^{-1}V_{Q_0}\psi_0. \quad (87)$$

Again, using the expansion Eq. 57, we find

$$\begin{aligned} \psi = \psi_0 + \lambda \sum_{k \neq 0} \psi_k \frac{V_{k0}}{E - E_k} \\ + \lambda^2 \sum_{k, h \neq 0} \psi_k \frac{V_{kh}}{E - E_k} \frac{V_{h0}}{E - E_h} + \dots \end{aligned} \quad (88)$$

As for the energy, if we arrest this expression to order  $n$  and substitute for  $E$  the value calculated by using Eq. 86, the result will differ from the one of Rayleigh-Schrödinger perturbation theory by terms of order  $n + 1$ .

A major drawback of the Brillouin-Wigner method is its lack of size consistency: for a system consisting of noninteracting subsystems, the perturbative correction to the energy of the total system is not the sum of the perturbative corrections to the energies of the separate subsystems through any finite order. This is best illustrated by the simple case of two systems  $a, b$  with unperturbed eigenvectors, energies, and interactions  $\psi_0^a, E_0^a$ , and  $\lambda V^a$ , and  $\psi_0^b, E_0^b$ , and  $\lambda V^b$ , respectively. If, for example, the expansion Eq. 86 is arrested at order 2, by noting that the matrix elements  $V_{0,ij}$  between the unperturbed state and the states  $\psi_i^a \psi_j^b$  are

$$\begin{aligned} V_{0,ij} &\equiv (\psi_0^a \psi_0^b, (V^a + V^b) \psi_i^a \psi_j^b) \\ &= (\psi_0^a, V^a \psi_i^a) \delta_{0j} + (\psi_0^b, V^b \psi_j^b) \delta_{0i}, \end{aligned}$$

for the second-order equation defining  $E$ , we find

$$\begin{aligned} E = E_0^a + E_0^b + \lambda \epsilon_1^a + \lambda \epsilon_1^b + \lambda^2 \sum_j \frac{|V_{0j}^b|^2}{E - E_0^a - E_j^b} \\ + \lambda^2 \sum_i \frac{|V_{0i}^a|^2}{E - E_0^b - E_i^a}. \end{aligned} \quad (89)$$

On the other hand, for the energy of each system at second order, we find

$$\begin{aligned} E^a &= E_0^a + \lambda \epsilon_1^a + \lambda^2 \sum_i \frac{|V_{0i}^a|^2}{E_a - E_i^a}; \\ E^b &= E_0^b + \lambda \epsilon_1^b + \lambda^2 \sum_j \frac{|V_{0j}^b|^2}{E_b - E_j^b}. \end{aligned} \quad (90)$$

It is apparent that the sum of the expression reported in Eq. 90 does not equal the expression of the energy reported in Eq. 89. This pathology is absent in the RS perturbation theory, where for noninteracting systems,  $E(\lambda) = E^a(\lambda) + E^b(\lambda)$ , hence, for any  $j$ ,  $\epsilon_j = (1/j!) D^j E(\lambda)|_{\lambda=0} = \epsilon_j^a + \epsilon_j^b$ .

## Symmetry and Degeneracy

In section “Perturbation of Point Spectra: Degenerate Case,” we applied perturbation theory to the case of degenerate eigenvalues with special emphasis on the problem of the removal of the degeneracy at a suitable order of perturbation theory. The main problem is to know in advance whether the degeneracy can be removed completely or a residual degeneracy is to be expected. The answer is given by group theory (Weyl 1931; Wigner 1959; Hamermesh 1989).

The very existence of degenerate eigenvalues of a Hamiltonian  $H$  is intimately connected with the symmetry properties of this operator. Generally speaking, a group  $G$  is a symmetry group for a physical system if there exists an associated set  $\{T(g)\}$  of transformations in the Hilbert space of the system such that  $|(T(g)\varphi, T(g)\psi)|^2 = |(\varphi, \psi)|^2$ ,  $g \in G$  (Wigner 1959). It is proven that the operators  $T(g)$  must be either unitary or antiunitary (Wigner 1959; Bargmann 1964). We will consider the most common case that they are unitary and can be chosen in such a way that

$$T(g_1)T(g_2) = T(g_1g_2), \quad g_1, g_2 \in G \quad (91)$$

so that the operators  $\{T(g)\}$  are a representation of  $G$ .

A system described by a Hamiltonian  $H$  is said to be invariant under the group  $G$  if the time evolution operator commutes with  $T(g)$ . Under fairly wide hypotheses, this implies

$$[H, T(g)] = 0, \quad g \in G. \quad (92)$$

A consequence is that, for any  $g \in G$ ,

$$H\psi = E\psi \Rightarrow HT(g)\psi = ET(g)\psi, \quad (93)$$

that is, the restrictions  $T(g)|_W$  of the operators  $T(g)$  to the space  $W$  corresponding to a given energy  $E$  are a representation of  $G$ . Given an orthonormal basis  $\{\psi_i\}$  in  $W$ , we have

$$T(g)\psi_i = \sum_j t_{ji}(g)\psi_j \quad (94)$$

and the vectors  $\psi_i$  are said to transform according to the representation of  $G$  described by the matrices  $t_{ji}$ .

This representation, apart from the occurrence of the so-called accidental degeneracy (which in most cases actually is a consequence of the invariance of the Hamiltonian under additional transformations), is irreducible: no subspace of  $W$  is invariant under all the transformations of the group. As a consequence, knowing the dimensions  $d_j$  of the irreducible representations of  $G$  allows to predict the possible degree of degeneracy of a given energy level, since the dimension of  $W$  must be equal to one of the numbers  $d_j$ . If the group of invariance is Abelian, all the irreducible representations are one-dimensional, and degeneracy can only be accidental.

Two irreducible representations are equivalent if there are bases which transform with the same matrix  $t_{ji}(g)$ . Otherwise they are inequivalent. The following orthogonality theorems hold. If  $a$  and  $b$  are inequivalent representations and  $\psi_i^{(a)}, \varphi_j^{(b)}$  transform according to these representations, then

$$\left(\psi_i^{(a)}, \varphi_j^{(b)}\right) = 0, \quad (95)$$

while, if  $a_r$  and  $a_s$  are equivalent, for the basis vectors  $\psi_i^{(a_r)}, \varphi_j^{(a_s)}$  we have

$$\left(\psi_i^{(a_r)}, \varphi_j^{(a_s)}\right) = K_{rs}^{(a)} \delta_{ij}. \quad (96)$$

Moreover, if  $A_{rs}$  is a matrix which commutes with all the matrices  $t_{ij}^{(a)}$  of an irreducible representation  $b$ , then  $A_{rs} = a\delta_{rs}$  (Schur's lemma).

### Symmetry and Perturbation Theory

If  $H = H_0 + \lambda V$ , let  $G_0$  be the group under which  $H_0$  is invariant. Although it is not the commonest case, we start with assuming that also the perturbation  $V$  commutes with  $T(g)$  for any element  $g$  of  $G_0$ . As a rule  $W_0$ , the space of eigenvectors of  $H_0$  with energy  $E_0$ , hosts an irreducible representation  $T(g)$  of  $G_0$ . In this case, the degeneracy cannot be removed at any order of perturbation theory. While this follows from general principles (for any value of  $\lambda$ ,  $\psi(\lambda)$  and  $T(g)\psi(\lambda)$  are eigenvectors of  $H(\lambda)$ , and by continuity the eigenspace  $W_\lambda$  will have the same dimension as  $W_0$ ), it is interesting to understand how the symmetry properties affect the mechanism of perturbation theory.

If  $\{\psi_i^0\}$  is a basis of  $W_0$  transforming according to an irreducible representation  $a$  of  $G_0$ , then the matrix  $V_{ij} = (\psi_i^0, V\psi_j^0)$  commutes with all the matrices  $t_{ji}^{(a)}(g)$  and, according to Schur's lemma,  $(\psi_i^0, V\psi_j^0) = v\delta_{ij} = \epsilon_1\delta_{ij}$ . No splitting occurs at the level of first-order perturbation theory, neither can it occur at any higher order. Indeed, when  $V$  commutes with the operators  $T(g)$ , then Bloch's operator  $U$  and consequently the operator  $B(\lambda)$  of Eq. 74, both commute with the  $T(g)$ 's too. Again by Schur's lemma, the operator  $B(\lambda)$  is a multiple of the identity. At any order of perturbation, the degeneracy of the level is not removed.

In most of the cases, however, the perturbation  $V$  does not commute with all the operators  $T(g)$ . The set

$$G = \{g : g \in G_0, [T(g), V] = 0\}$$

is a subgroup  $G$  of  $G_0$ , and the group of invariance for the Hamiltonian  $H$  is reduced to  $G$ .  $W_0$  generally contains  $G$ -irreducible subspaces  $W_i$ ,  $1 \leq i \leq n$ : the operators  $T(g)|_{W_0}$  are a reducible representation of  $G$ . The decomposition into irreducible representations of  $G$  is unique up to equivalence.

The crucial information we gain from group theory is the following: the number of energy levels which the energy  $E_0$  is split into is the number of irreducible representations of  $G$  which the representation of  $G_0$  in  $W_0$  is split into. The degrees of degeneracy are the dimensions of these representations. What is relevant is that we only need to study the eigenspace  $W_0$  of  $H_0$ , which is known by hypothesis.

In fact, let  $W(\lambda)$  be the space spanned by the eigenvectors  $\psi_k(\lambda)$  of  $H(\lambda)$  such that  $\psi_k(0) \in W_0$ .  $W(\lambda)$  is invariant under the operators  $T(g)$ ,  $g \in G$ , since Bloch's operator  $U$  commutes with the operators  $T(g)$ ,  $g \in G$ .  $W(\lambda)$  can be decomposed into  $G$ -irreducible subspaces  $W_k(\lambda)$ , and in each of them by Schur's lemma; the Hamiltonian  $H(\lambda)$  is represented by a matrix  $E_k(\lambda)I_{W_k(\lambda)}$ . The projections  $P_0 W_k(\lambda)$  span the space  $W_0$  and transform with the same representation of  $G$  as  $W_k(\lambda)$ , since  $P_0$  commutes with  $T(g)$  for any  $g$  in  $G_0$ , hence for any  $g$  in  $G$ . Thus, the space  $W_0$  hosts as many irreducible representations of  $G$  as  $W(\lambda)$ . Assuming that the eigenvalues  $E_k(\lambda)$  are different from one another, we see that the decomposition of the representation of  $G_0$  in  $W_0$  into irreducible representations of  $G$  determines the number and the degeneracy of the eigenvalues of  $H(\lambda)$  such that the corresponding eigenvectors  $\psi(\lambda)$  are in  $W_0$  for  $\lambda = 0$ . The possibility that some of the  $E_k(\lambda)$  are equal will be touched upon in the next subsection.

Examples where the above mechanism is at work are common in atomic physics. When an atom, whose unperturbed Hamiltonian  $H_0$  is invariant under  $O(3)$ , is subjected to a constant electric field  $\vec{E} = E\hat{z}$  (Stark effect), the invariance group  $G$  of its Hamiltonian is reduced to the rotations about the  $z$  axis ( $SO(2)$ ) and the reflections with respect to planes containing the  $z$  axis. The irreducible representations of this group have dimension at most 2, and the  $G$ -irreducible subspaces of  $W_0$  (the space generated by the eigenvectors  $\psi_{E_0lm}$  of  $H_0$  corresponding to the energy  $E_0$ ) are generated by  $\psi_{E_0/0}$  (one-dimensional representation) and  $\psi_{E_0/\pm m}$  (two-dimensional representations). Hence, the level  $E_0$  is split into  $l + 1$  levels, the states with  $m$  and  $-m$  remaining degenerate since reflections transform a vector

with a given  $m$  into the vector with opposite  $m$ . Instead, if the atom is subjected to a constant magnetic field  $\vec{B} = B\hat{z}$  (Zeeman effect), the surviving invariance group  $G$  consists of  $SO(2)$  plus the reflections with respect to planes  $z = z_0$ .  $G$  being Abelian, the degeneracy is completely removed, and this occurs at the first order of perturbation theory.

In the rather special case that  $W_0$  contains subspaces transforming according to inequivalent representations of  $G_0$ , also a  $G_0$ -invariant perturbation  $V$  can separate in energy the states belonging to inequivalent representations. For example, the spectrum of alkali atoms can be calculated by considering in a first approximation an electron in the field of the unit-charged atomic core, which is treated as pointlike. In this problem, the obvious invariance group of the Hamiltonian of the optical electron is  $O(3)$ , the group of rotations and reflections, and the space  $W_0$  corresponding to the principal quantum number  $n > 1$  contains  $n$  inequivalent irreducible representations which are labeled by the angular momentum  $l \leq n - 1$ . When the finite dimension of the atomic rest is taken into account as a perturbation, its invariance under  $O(3)$  splits the levels with given  $n$  and different  $l$  into  $n$  sublevels. A more careful consideration, however, shows that also the Lenz vector commutes with the unperturbed Hamiltonian (Böhm 1993) and that the space  $W_0$  is irreducible under a larger group, the group  $SO(4)$  (Fock 1935), which is generated by the angular momentum and the Lenz vector. As a consequence, the  $l$  degeneracy is by no means accidental: a space irreducible under a given group can turn out to be reducible with respect to one of its subgroups.

Group theory is a valuable tool in degenerate perturbation theory to search the correct vectors  $\psi_k(0)$  which make the operator  $P_0VP_0$  diagonal. In fact, let  $\psi_i^{(a)}$  be vectors which reduce the representation  $T$  of  $G$  in  $W_0$  into its irreducible components  $T^{(a)}$ . The vectors  $\psi_i^{(a)}$  and  $V\psi_i^{(a)}$  transform according to the same irreducible representation  $T^{(a)}$ . Hence, by Eqs. 95 and 96, we find that the  $P_0VP_0$  is a diagonal block matrix with respect to inequivalent representations:



$$\left(\psi_i^{(a_r)}, V\psi_j^{(b_s)}\right) = K_{rs}^{(a)} \delta_{ij} \delta_{ab}, \quad (97)$$

with  $\delta_{ab} = 1$  if representations  $a$  and  $b$  are equivalent,  $\delta_{ab} = 0$  otherwise. The matrices  $K_{rs}^{(a)}$  are generally much smaller than the full matrix of the potential. Thus, the operation of diagonalizing  $V$  is made easier by finding the  $G$ -irreducible subspaces  $W_a$ . Conversely, the reduction of an irreducible representation of a group  $G_0$  in a space  $W_0$  into irreducible representations of a subgroup  $G$  can be achieved by the following trick: find an operator  $V$  whose symmetry group is just  $G$  and interpret  $W_0$  as the degeneracy eigenspace of a Hamiltonian  $H_0$ . The  $G$ -irreducible subspaces of  $W_0$  are the eigenspaces of  $P_0VP_0$ .

### Level Crossing

As shown in the foregoing section, the existence of a non-Abelian group of symmetry for the Hamiltonian entails the existence of degenerate eigenvalues. The problem naturally arises as to whether there are cases when, on the contrary, the degeneracy is truly “accidental,” that is, it cannot be traced back to symmetry properties. The problem was discussed by J. von Neumann and E.P. Wigner (1929), who showed that for a generic  $n \times n$  Hermitian matrix depending on real parameters  $\lambda_1, \lambda_2, \dots$ , three real values of the parameters have to be adjusted in order to have the collapse of two eigenvalues (level crossing).

When passing to infinite dimension, arguments valid for finite dimensional matrices might fail. Moreover, often the Hamiltonian is not sufficiently “generic” so that level crossing may occur. As a consequence, we look for necessary conditions in order that, given the Hamiltonian  $H(\lambda) = H_0 + \lambda V$ , two eigenvalues collapse for some (real) value  $\bar{\lambda}$  of the parameter  $\lambda$ :  $E_1(\bar{\lambda}) = E_2(\bar{\lambda}) \equiv \bar{E}$ . In this case, if  $\psi_1(\bar{\lambda})$  and  $\psi_2(\bar{\lambda})$  are any two orthonormal eigenvectors of  $H(\bar{\lambda}) = H_0 + \bar{\lambda}V$  belonging to the eigenvalue  $\bar{E}$ , the matrix

$$H_{ij}(\bar{\lambda}) \equiv \left(\psi_i(\bar{\lambda}), (H_0 + \bar{\lambda}V)\psi_j(\bar{\lambda})\right), i, j = 1, 2$$

must be a multiple of the identity:

$$H_{11}(\bar{\lambda}) = H_{22}(\bar{\lambda}) \quad (98)$$

$$H_{12}(\bar{\lambda}) = 0. \quad (99)$$

Equations 98 and 99 are three real equations for the unknown  $\bar{\lambda}$ ; hence, except for special cases, level crossing cannot occur.

The condition expressed by Eq. 99 is satisfied if the states corresponding to the eigenvalues  $E_1(\lambda)$  and  $E_2(\lambda)$  possess different symmetry properties, that is, if they belong to inequivalent representations of the invariance group of the Hamiltonian or, equivalently, if they are eigenvectors with different eigenvalues of an operator which for any  $\lambda$  commutes with the Hamiltonian  $H(\lambda)$  (hence, it commutes with both  $H_0$  and  $V$ ). In this case,  $H_{12} = 0$  and the occurrence of level crossing depends on whether Eq. 98 has a real solution. This explains the statement that level crossing can occur only for states with different symmetry, while states of equal symmetry repel each other. Indeed, if Eq. 99 is not satisfied, the behavior of two close eigenvalues as functions of  $\lambda$  is illustrated in Fig. 1 (Section “Presentation of the Problem and an Example”).

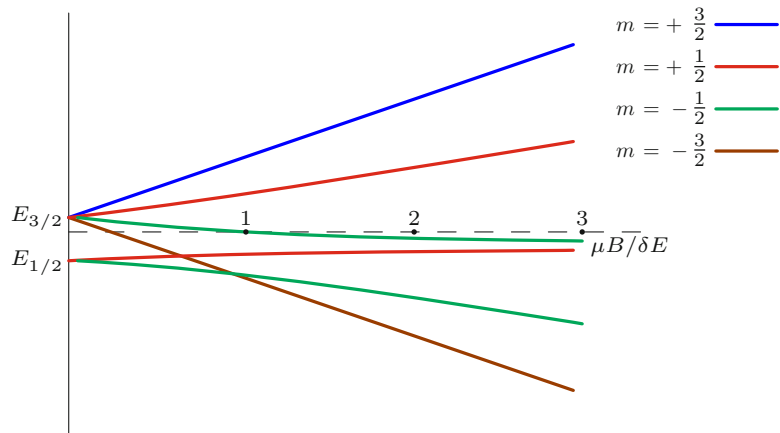
Figure 2 illustrates the behavior of the quasi-degenerate energy levels  $2p_{1/2}, 2p_{3/2}$  of the lithium atom in the presence of an external magnetic field  $\vec{B}$ . In the absence of the magnetic field, they are split by the spin-orbit interaction, with a separation  $\delta E \equiv E_{3/2} - E_{1/2} = 0.4 \times 10^{-4}$  eV, to be compared with the separation in excess of 1 eV from the adjacent  $2s$  and  $3s$  levels. This justifies treating the effect of the magnetic field by means of the first-order perturbation theory for quasi-degenerate levels.

When the magnetic field is present, the residual symmetry is the (Abelian) group of rotations about the direction of  $\vec{B}$ . Hence, the Hamiltonian commutes with the component of the angular momentum along the direction of  $\vec{B}$ , whose eigenvalues are denoted with  $m$ . In Fig. 2, the energies of states with equal symmetry, that is, with the same value of  $m$ , are depicted with the same color. No crossing occurs between states with equal  $m$ , while the level with  $m = -3/2$  does cross both the



### Perturbation Theory in Quantum Mechanics,

**Fig. 2** The effect of a magnetic field on the doublet  $2p_{1/2}$ ,  $2p_{3/2}$  of the lithium whose degeneracies, in the absence of the magnetic field, are, respectively, 2 and 4.  $\mu$  is the magnetic moment of the electron and  $\mu B/\delta E = 1$  for  $B \approx 1.4$  T



levels with  $m = 1/2$  and with  $m = -1/2$  which the  $2p_{1/2}$  level is split into.

### Problems with the Perturbation Series

So far, we have assumed that all the power expansions appearing in the calculations were converging for  $|\lambda| \leq 1$ , that is, we assumed analyticity in  $\lambda$  of  $E(\lambda)$ . Actually, it is only for rather special cases that analyticity can be proved. For most of the cases of physical interest, even if the terms of the perturbation series can be shown to exist, the series does not converge or, when it converges, the limit is not  $E(\lambda)$ . In spite of this, special techniques have been devised to extract a good approximation to  $E(\lambda)$  from the (generally few) terms of the perturbation series which can be computed. We will outline the main results existing in the field, without delving into mathematical details, for which we refer the reader to the books of Kato (1966) and Reed–Simon (1978) and the references therein.

The most favorable case is that of the so-called regular perturbations (Rellich 1937a, b, 1939, 1940), where the perturbation series does converge to  $E(\lambda)$ . More precisely, if  $E_0$  is a non-degenerate eigenvalue of  $H_0$ , for  $\lambda$  in a suitable neighborhood of  $\lambda = 0$ , the Hamiltonian  $H = H_0 + \lambda V$  has a nondegenerate eigenvalue  $E(\lambda)$  which is analytic in  $\lambda$  and equals  $E_0$  for  $\lambda = 0$ . The same property holds for the eigenvector  $\psi(\lambda)$ . A sufficient condition for this property to hold is

expressed by the Kato-Rellich theorem (Rellich 1937a, b, 1939, 1940; Kato 1949), which essentially states that if the perturbation  $V$  is  $H_0$ -bounded, in the sense that constants  $a, b$  exist such that

$$\|V\psi\| \leq a\|H_0\psi\| + b\|\psi\| \quad (100)$$

for any  $\psi$  in the domain of  $V$  (which must include the domain of  $H_0$ ), then the perturbation is regular. A lower bound to the radius  $r$  such that the perturbation series converges to the eigenvalue  $E(\lambda)$  for  $|\lambda| < r$  can be given in terms of the parameters  $a, b$  appearing in Eq. 100 and the distance  $\delta$  of the eigenvalue  $E_0$  from the rest of the spectrum of  $H_0$ . We have

$$r = \left[ a + \frac{2}{\delta} \left[ b + a \left( |E_0| + \frac{\delta}{2} \right) \right] \right]^{-1}. \quad (101)$$

It must be stressed, however, that the constants  $a$  and  $b$  are not uniquely determined by  $V$  and  $H_0$ . If the perturbation  $V$  is bounded ( $a = 0, b = \|V\|$ ) condition, Eq. 101 reads  $r = \delta/(2\|V\|)$ , which implies that the perturbation series for  $H = H_0 + V$  with  $V$  bounded converges if  $\|V\| < \delta/2$  (Kato bound (Kato 1949)). The analysis of the two-level system (Section “Presentation of the Problem and an Example”) shows that the figure 1/2 cannot be improved. Still, Kato bound is only a lower bound to  $r$ .

A similar statement holds for degenerate eigenvalues (Reed and Simon 1978): if  $E_0$  has

multiplicity  $m$ , there are  $m$  single valued analytic functions  $E_k(\lambda)$ ,  $k = 1, \dots, m$  such that  $E_k(0) = E_0$  and, for  $\lambda$  in a neighborhood of 0,  $E_k(\lambda)$  are eigenvalues of  $H(\lambda) = H_0 + \lambda V$ . Some of the functions  $E_k(\lambda)$  may be coincident, and in a neighborhood of  $E_0$ , there are no other eigenvalues of  $H(\lambda)$ .

Regular perturbations are in fact exceedingly rare, a notable case being that of helium-like atoms (Simon 1991). Actually, there are cases where, although on physical grounds  $H_0 + \lambda V$  does possess bound states, the relationship between  $E(\lambda)$  and the RS expansion is far more complicated than for regular perturbations. As pointed out by Kramers (1957), with an argument similar to an observation by Dyson (1952) for quantum electrodynamics, the quartic anharmonic oscillator with Hamiltonian

$$H = H_0 + \lambda V \\ \equiv \frac{p^2}{2m} + \frac{m\omega^2 x^2}{2} + \lambda \frac{m^2 \omega^3}{\hbar} x^4 \quad (102)$$

is such an example. In fact, on the one hand, bound states exist only for  $\lambda \geq 0$ ; on the other hand, if a power series converges for  $\lambda > 0$ , then the series should converge also for negative values of  $\lambda$ . But for  $\lambda < 0$ , no bound state exists. Still worse, by estimating the coefficients of the RS expansion, it has been proved that the series has vanishing radius of convergence (Bender and Wu 1969).

In spite of this negative result, in this case it has been proved (Simon 1970) that the perturbation series is an asymptotic series. This means that, for each  $n$ , if  $\sum_0^n \epsilon_k \lambda^k$  is the sum through order  $n$  of the perturbation series, then

$$\lim_{\lambda \rightarrow 0} \frac{\sum_0^n \epsilon_k \lambda^k - E(\lambda)}{\lambda^n} = 0. \quad (103)$$

We recall the difference between an asymptotic and an absolutely converging series, such as occurs with regular perturbations. For the latter one, given any  $\lambda$  in the convergence range of the series, the distance  $|\sum_0^n \epsilon_k \lambda^k - E(\lambda)|$  can be made arbitrarily small provided  $n$  is sufficiently large (so that a converging series is also an asymptotic

series). On the contrary, for an asymptotic series  $|\sum_0^n \epsilon_k \lambda^k - E(\lambda)|$  is arbitrarily small only if  $\lambda$  is sufficiently near 0, but for a definite value of  $\lambda$ , the quantity  $|\sum_0^n \epsilon_k \lambda^k - E(\lambda)|$  might decrease to a minimum, attained for some value  $N$ , and then it could start to oscillate for  $n > N$  (this is indeed the case for the anharmonic oscillator). As a consequence, for asymptotic series it is not expedient to push the calculation of the terms of the series beyond the limit where wild oscillations set in.

Any  $C^\infty$  function has an asymptotic series, as can be seen by inspection of the Taylor's formula with a remainder (see Eq. 44). By this means, Krieger (1968) argued that, if  $\epsilon_k$  (or equivalently the  $k$ -th derivative of  $E(\lambda)$ ) exists for any  $k$ , the RS series is asymptotic. However, generally there is not a range where the series converges to  $E(\lambda)$ , that is,  $E(\lambda)$  is not analytic. An asymptotic series may fail to converge at all for  $\lambda \neq 0$ , as noted for the anharmonic oscillator. The asymptotic series of a function, if it exists, is unique, but the converse is not true. For example, for the  $C^\infty$  function defined for real  $x$  as  $f(x) = \exp(-1/x^2)$  if  $x \neq 0$ ,  $f(0) = 0$ , the asymptotic series vanishes. There are also cases when the perturbation series is asymptotic for  $\arg \lambda$  lying in a range  $[\alpha, \beta]$ . This occurs, for example, for the generalized anharmonic oscillator with perturbation  $V \propto \lambda x^{2n}$ . It has been proved that its perturbation series is asymptotic for  $|\arg \lambda| \leq \theta < \pi$  (Simon 1970) (note that the domain does not include negative values of  $\lambda$ ). The result was later extended to multidimensional anharmonic oscillators (Graffi et al. 1970). General theorems stating sufficient hypotheses for the perturbation series to be asymptotic can be found in the literature. As a rule, however, they do not cover most of the cases of physical interest.

Even in the felicitous case when the perturbation series is asymptotic, it is only known that a partial sum approaches  $E(\lambda)$  as much as desired provided  $\lambda$  is sufficiently small. This is not of much help to the practicing scientist, who generally is confronted with a definite value of the parameter  $\lambda$ , which can always be considered  $\lambda = 1$  by an appropriate rescaling of the potential  $V$ . Recalling that different functions can have the same asymptotic series, it seems hopeless to try to recover the function  $E(\lambda)$  from its asymptotic

series, but this is possible for the so-called strong asymptotic series. A function  $E(\lambda)$  analytic in a sectorial region ( $0 < |\lambda| < B$ ,  $|\arg \lambda| < \pi/2 + \delta$ ) is said to have strong asymptotic series  $\sum_0^\infty a_k \lambda^k$  if for all  $\lambda$  in the sector

$$\left| E(\lambda) - \sum_0^n a_k \lambda^k \right| \leq C \sigma^{n+1} |\lambda|^{n+1} (n+1)! \quad (104)$$

for some constants  $C, \sigma$ . For strong asymptotic series, it is proved that the function  $E(\lambda)$  is uniquely determined by the series. Conditions that ensure that the RS series is a strong asymptotic series have been given (Reed and Simon 1978).

The problem of actually recovering the function  $E(\lambda)$  from its asymptotic series can be tackled by several methods. The most widely used procedure is the Borel summation method (Borel 1899), which amounts to what follows. Given the strongly asymptotic series  $\sum_0^\infty a_k \lambda^k$ , one considers the series  $F(\lambda) \equiv \sum_0^\infty (a_k/k!) \lambda^k$ . This is known as the Borel transform of the initial series, which, by the hypothesis of strong asymptotic convergence, can be proved to have a nonvanishing radius of convergence and to possess an analytic continuation to the positive real axis. Then, the function  $E(\lambda)$  is given by

$$E(\lambda) = \int_0^\infty F(\lambda x) \exp(-x) dx. \quad (105)$$

The above statement is Watson's theorem (Watson 1912). Roughly speaking, it yields the function  $E(\lambda)$  as if the following exchange of the series with the integral were allowed:

$$E(\lambda) \sim \sum_0^\infty a_k \lambda^k = \sum_0^\infty \frac{a_k}{k!} \int_0^\infty \exp(-x) x^k dx \lambda^k = \int_0^\infty \exp(-x) \sum_0^\infty \frac{a_k}{k!} (x\lambda)^k dx = \int_0^\infty F(\lambda x) \exp(-x) dx.$$

A practical problem with perturbation theory is that, apart from a few classroom examples, one is able to calculate only the lower-order terms of the perturbation series. Although in principle it is impossible to divine the rest of a series by

knowing its terms through a given order, a technique which in some cases turned out to work is that of Padé approximants (Padé 1899; Baker and Graves-Morris 1996). A Padé  $[M, N]$  approximant to a series is a rational function:

$$R_{MN}(z) = \frac{P_M(z)}{Q_N(z)} \quad (106)$$

whose power expansion near  $z = 0$  is equal to the first  $M + N$  terms of the series. It has been proved (Loeffel et al. 1969) that the Padé  $[N, N]$  approximants converge to the true eigenvalue of the anharmonic oscillator with  $x^4$  or  $x^6$  perturbation. The Padé  $[M, N]$  approximant to a function  $f(z)$  is unique, but its domain of analyticity is generally larger. Even for asymptotic series whose first terms are known, one can write the Padé approximants. One can either use directly the Padé approximant as the value of  $E(\lambda)$  for the desired value of  $\lambda$  or can insert it into the Borel summation method. For the case of the quartic anharmonic oscillator (Eq. 102), both methods have been proved to work (at the cost of calculating some tens of terms of the series).

Another approach to the problem is the method of self-similar approximants (Yukalov and Yukalova 2007), whereby approximants to the function  $E(\lambda)$  for which the asymptotic series is known are sought by means of products

$$f_{2p}(\lambda) = \prod_{i=1}^p (1 + A_i \lambda)^{n_i}. \quad (107)$$

The  $2p$  parameters  $A_i, n_i$ ,  $1 \leq i \leq p$  are determined by equating the Taylor expansion of  $f_{2p}(\lambda)$  with the asymptotic series through order  $2p$  ( $a_0 = 1$  can be assumed, with no loss of generality; see Yukalov and Yukalova 2007). Also, odd-order approximants  $f_{2p+1}$  are possible. For the anharmonic oscillator (Eq. 102), the calculations exhibit a steady convergence to the correct value of the energy of both the even-order and the odd-order approximants also for  $\lambda = 200$ .

The problem with the above approaches is that their efficiency seems limited to toy models as the anharmonic oscillator. For realistic problems, it is difficult to establish in advance that the method converges to the correct answer.

## Perturbation of the Continuous Spectrum

In this section, we consider the effect of a perturbing potential  $V$  on states belonging to the continuous spectrum. Since the problem is interesting mainly for the theory of scattering, we will assume that the unperturbed Hamiltonian  $H_0$  is the free Hamiltonian of a particle of mass  $m$ . Also, assuming that the potential  $V(\vec{r})$  vanishes at infinity, the spectrum of the free Hamiltonian  $H_0$  and the continuous spectrum of the exact Hamiltonian  $H = H_0 + \lambda V$  are equal and consist of the positive real semiaxis. Given an energy  $E = \hbar^2 k^2 / 2m$ , the problem is how the potential  $V$  affects that particular eigenfunction  $\psi_0$  of  $H_0$  which would represent the state of the system if the interaction potential were absent.

Letting  $\psi = \psi_0 + \delta\psi$ , the Schrödinger equation reads

$$(E - H_0)\delta\psi = \lambda V(\psi_0 + \delta\psi). \quad (108)$$

In the spirit of the perturbation approach,  $\delta\psi$  can be calculated by an iterative process provided we are able to find the solution of the inhomogeneous equation

$$(E - H_0)\delta\psi = \zeta \quad (109)$$

in the form

$$\delta\psi = \tilde{G}_0 \zeta, \quad (110)$$

$\tilde{G}_0$  being the Green function of Eq. 109. Assuming that  $\tilde{G}_0$  is known, we find

$$\begin{aligned} \delta\psi &= \lambda \tilde{G}_0 V \psi = \lambda \tilde{G}_0 V \psi_0 + \lambda \tilde{G}_0 V \delta\psi \\ &= \lambda \tilde{G}_0 V \psi_0 \\ &\quad + \lambda \tilde{G}_0 V (\lambda \tilde{G}_0 V \psi_0 + \lambda \tilde{G}_0 V \delta\psi) \\ &= \lambda \tilde{G}_0 V \psi_0 + \lambda^2 \tilde{G}_0 V \tilde{G}_0 V \psi_0 \\ &\quad + \lambda^2 \tilde{G}_0 V \tilde{G}_0 V \delta\psi \\ &= \dots, \end{aligned} \quad (111)$$

that is,  $\delta\psi$  is written as a power expansion in  $\lambda$  in terms of the free wave function  $\psi_0$ .

## Scattering Solutions and Scattering Amplitude

One has to decide which eigenfunction of  $H_0$  must be inserted into the above expression and which Green function  $\tilde{G}_0$  must be used, since, of course, the solution of Eq. 109 is not unique. The questions are strongly interrelated, and the answers depend on which solution of the exact Schrödinger equation one wishes to find. Since the study of the perturbation of the continuous spectrum is relevant mainly for the theory of potential scattering, we will focus on this aspect. In the theory of scattering, it is shown (Joachain 1983) that, for a potential  $V(\vec{r})$  vanishing faster than  $1/r$  for  $r \rightarrow \infty$ , a wave function  $\psi$  which in the asymptotic region is an eigenfunction of the momentum operator plus an outgoing wave

$$\psi \xrightarrow{r \rightarrow \infty} \exp(i\vec{k} \cdot \vec{r}) + f_{\vec{k}}(\theta, \varphi) \frac{\exp(ikr)}{r} \quad (112)$$

( $\theta, \varphi$  being the polar angles with respect to the  $\vec{k}$  axis) is suitable for describing the process of diffusion of a beam of free particles with momentum  $\vec{k}$  which impinge onto the interaction region and are scattered according the amplitude  $f_{\vec{k}}(\theta, \varphi)$ . (The character of outgoing wave of the second term in Eq. 112 is apparent when the time factor  $\exp(-iEt)$  is taken into account.)

The differential cross section  $d\sigma/d\Omega$  is the ratio of the flux of the probability current density due to the outgoing wave to the flux due to the impinging plane wave. One finds

$$\frac{d\sigma}{d\Omega} = |f_{\vec{k}}(\theta, \varphi)|^2. \quad (113)$$

In conclusion, we require that  $\psi_0$  is a plane wave and that the Green function  $\tilde{G}_0$  has to be chosen in such a way as to yield an outgoing wave for large  $r$ .

Thus, we need to solve the equation

$$\begin{aligned} \left(\vec{k}^2 + \Delta\right)\delta\psi &= \lambda \frac{2m}{\hbar^2} V \left( \exp(i\vec{k} \cdot \vec{r}) + \delta\psi \right) \\ &\equiv U(\vec{r}) \left( \exp(i\vec{k} \cdot \vec{r}) + \delta\psi \right) \end{aligned} \quad (114)$$

with the asymptotic condition  $\delta\psi \rightarrow f_{\vec{k}}^- \exp(ikr)/r$  for  $r \rightarrow \infty$ . In terms of the Green function  $G_0(\vec{r}, \vec{r}')$ , which satisfies the equation

$$(\Delta + k^2)G_0(\vec{r}, \vec{r}') = \delta(\vec{r} - \vec{r}'), \quad (115)$$

the solution of Eq. 114 can be written as

$$\delta\psi(\vec{r}) = \int G_0(\vec{r}, \vec{r}') U(\vec{r}') \left[ \left( \exp(i\vec{k} \cdot \vec{r}') \right) + \delta\psi(\vec{r}') \right] d\vec{r}', \quad (116)$$

which is a form of the Lippmann-Schwinger equation (Lippmann and Schwinger 1950). The integral operator  $G_0$  with kernel  $G_0(\vec{r}, \vec{r}')$  is connected to the operator  $\tilde{G}_0$  of Eq. 110 by the equation  $\tilde{G}_0 = 2mG_0/\hbar^2$ .

The leading term of  $\delta\psi$  for  $r \rightarrow \infty$  is determined by the leading term of  $G_0(\vec{r}, \vec{r}')$ , so we look for a solution of Eq. 115 with the behavior of outgoing wave for  $r \rightarrow \infty$ . Due to translation and rotation invariance (if both the incoming beam and the scattering potential are translated or rotated by the same amount, the scattering amplitude  $f_{\vec{k}}^-(\theta, \varphi)$  is unchanged), we require for the solution a dependence only on  $|\vec{r} - \vec{r}'|$ .

Recalling that  $\Delta 1/r = -4\pi\delta(\vec{r})$ , we look for a solution of Eq. 115 with  $\vec{r}' = 0$  of the form  $-F(r)/(4\pi r)$ , with  $F(0) = 1$ . The function  $G_0(\vec{r}, \vec{r}')$  then will be

$$G_0(\vec{r}, \vec{r}') = \frac{F(|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|}. \quad (117)$$

The equation for  $F(r)$  is

$$F'' + k^2 F = 0, \quad (118)$$

whose solutions are  $\exp(\pm ikr)$  (outgoing and incoming wave, respectively). In conclusion for  $G_0$ , we find

$$G_0(\vec{r}, \vec{r}') = -\frac{1}{4\pi} \frac{\exp(ik|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|}. \quad (119)$$

The solution of the Schrödinger equation with the Green function given in Eq. 119 is denoted as  $\psi_{\vec{k}}^+$  and obeys the integral equation known as the Lippmann-Schwinger equation (Lippmann and Schwinger 1950):

$$\psi_{\vec{k}}^+(\vec{r}) = \exp(i\vec{k} \cdot \vec{r}) - \frac{1}{4\pi} \int \frac{\exp(ik|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|} U(\vec{r}') \psi_{\vec{k}}^+(\vec{r}') d\vec{r}'. \quad (120)$$

The behavior for  $r \rightarrow \infty$  can be easily checked to be as in Eq. 112 by inserting the expansion

$$|\vec{r} - \vec{r}'| = r - \frac{\vec{r} \cdot \vec{r}'}{r} + O(1/r) \quad (121)$$

into the Green function  $G_0$ . We find  $(\hat{r} \equiv \vec{r}/r)$

$$-\frac{1}{4\pi} \frac{\exp(ik|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|} \xrightarrow{r \rightarrow \infty} -\frac{1}{4\pi} \frac{\exp[ik(r - \hat{r} \cdot \vec{r}')] }{r} \left[ 1 + \frac{\hat{r} \cdot \vec{r}'}{r^2} \right], \quad (122)$$

which yields for  $\psi_{\vec{k}}^+(\vec{r})$  ( $\vec{k}_f \equiv k\hat{r}$ )

$$\psi_{\vec{k}}^+(\vec{r}) \xrightarrow{r \rightarrow \infty} \exp(i\vec{k} \cdot \vec{r}) - \frac{1}{4\pi} \frac{\exp(ikr)}{r} \int \exp(-i\vec{k}_f \cdot \vec{r}') U(\vec{r}') \psi_{\vec{k}}^+(\vec{r}') d\vec{r}'. \quad (123)$$

The solutions  $\psi_{\vec{k}}^+(\vec{r})$  are normalized as the plane waves  $\exp(i\vec{k} \cdot \vec{r})$ :

$$(\psi_{\vec{k}}^+, \psi_{\vec{k}}^+) = (2\pi)^3 \delta(\vec{k} - \vec{k}'). \quad (124)$$

In addition, they are orthogonal to any possible bound state solution of the Schrödinger equation with the Hamiltonian  $H = H_0 + \lambda V$ . Together

with the bound state solutions, they constitute a complete set. On a par with the solutions  $\psi_k^+(\vec{r})$ , one can also envisage solutions  $\psi_k^-(\vec{r})$  with asymptotic behavior of incoming wave. They are obtained using for  $F$  (see Eq. 118) the solution  $\exp(-ikr)$ . The normalization and orthogonality properties of the functions  $\psi_k^-(\vec{r})$  are the same as for the  $\psi_k^+(\vec{r})$  functions.

From Eq. 123, we derive an implicit expression for the scattering amplitude  $f_k^-(\theta, \varphi)$ :

$$f_k^-(\theta, \varphi) = -\frac{1}{4\pi} \int \exp(-i\vec{k}_f \cdot \vec{r}') U(\vec{r}') \psi_k^+(\vec{r}') d\vec{r}' \quad (125)$$

where the unknown function  $\psi_k^+(\vec{r}')$  still appears. Letting  $\varphi_f \equiv \exp(i\vec{k}_f \cdot \vec{r}')$ , Eq. 125 can also be written as

$$f_k^-(\theta, \varphi) = -\frac{1}{4\pi} (\varphi_f, U\psi_k^+). \quad (126)$$

### The Born Series and Its Convergence

Equations 120 and 126 are the starting point to obtain the expression of the exact wave function  $\psi_k^+(\vec{r})$  and the scattering amplitude  $f_k^-(\theta, \varphi)$  as a power series in  $\lambda$ , in the spirit of the perturbation approach. Recalling that  $U = (2m/\hbar^2)V$  (see Eq. 114), if  $\varphi_k^- \equiv \exp(i\vec{k} \cdot \vec{r})$  for  $\psi_k^+(\vec{r})$  we find

$$\begin{aligned} \psi_k^+ &= \varphi_k^- + \lambda G_0 \frac{2m}{\hbar^2} V \varphi_k^- + \lambda^2 G_0 \frac{2m}{\hbar^2} V G_0 \frac{2m}{\hbar^2} V \varphi_k^- \\ &+ \dots = \exp(i\vec{k} \cdot \vec{r}) + \lambda \int G_0(\vec{r}, \vec{r}') \frac{2m}{\hbar^2} V(\vec{r}') \\ &\exp(i\vec{k} \cdot \vec{r}') d\vec{r}' + \lambda^2 \int d\vec{r}' \int d\vec{r}'' G_0(\vec{r}, \vec{r}') \frac{2m}{\hbar^2} \\ &V(\vec{r}') G_0(\vec{r}', \vec{r}'') \frac{2m}{\hbar^2} V(\vec{r}'') \exp(i\vec{k} \cdot \vec{r}'') \\ &+ \dots \end{aligned} \quad (127)$$

Inserting the above expansion into Eq. 126, for the scattering amplitude  $f_k^-(\theta, \varphi)$ , we find

$$f_k^-(\theta, \varphi) = \sum_{n=1}^{\infty} f_k^{(n)}(\theta, \varphi), \quad (128)$$

where  $f_k^{(n)}$ , the contribution of order  $n$  in  $\lambda$  to  $f_k^-(\theta, \varphi)$ , is obtained by substituting  $\psi_k^+$  in Eq. 126 with the contribution of order  $n-1$  of the expansion Eq. 127. The term of order 1 is called the Born approximation (Born 1926) and is given by

$$\begin{aligned} f_k^B(\theta, \varphi) &\equiv f_k^1(\theta, \varphi) \\ &= -\frac{1}{4\pi} \left( \frac{2m\lambda}{\hbar^2} \right) \int \exp[i(\vec{k} - \vec{k}_f) \cdot \vec{r}'] \\ &V(\vec{r}') d\vec{r}'. \end{aligned} \quad (129)$$

The term of order 2 is

$$f_k^2(\theta, \varphi) = -\frac{1}{4\pi} \left( \frac{2m\lambda}{\hbar^2} \right)^2 (\varphi_f, V G_0 V \varphi_k^-), \quad (130)$$

and the general term of order  $n$  is

$$\begin{aligned} f_k^{(n)}(\theta, \varphi) &= -\frac{1}{4\pi} \left( \frac{2m\lambda}{\hbar^2} \right)^n \\ &(\varphi_f, V G_0 V \dots G_0 V \varphi_k^-) (n \text{ times } V). \end{aligned} \quad (131)$$

The scattering amplitude through order  $n$  is

$$f_k^{[n]}(\theta, \varphi) = \sum_{i=1}^n f_k^{(i)}(\theta, \varphi) \quad (132)$$

with  $f_k^{(1)}(\theta, \varphi) \equiv f_k^B(\theta, \varphi)$ , and the series  $\sum_0^\infty f_k^{(i)}(\theta, \varphi)$  is known as the Born series (Born 1926). Of course, when using Eq. 132 for calculating the differential cross section  $d\sigma/d\Omega$ , only terms of order not exceeding  $n$  should be consistently retained.

For a discussion of the range of validity and the convergence of the expansions Eqs. 127 and 128, it is convenient to pose the problem in the framework of integral equations in the Hilbert space  $L^2$  (Riesz and Sz.-Nagy 1968; Yosida 1991), which provides a natural notion of convergence. To this purpose,

since  $\psi_k^+(\vec{r})$  is not square integrable, we start assuming that the potential  $V(\vec{r})$  is summable:

$$\int |V(\vec{r})| d\vec{r} < \infty \quad (133)$$

and multiply Eq. 120 by  $|V(\vec{r})|^{\frac{1}{2}}$  (Rollnik 1956; Grossman 1961; Schwartz 1960). Letting  $\epsilon_V(\vec{r}) \equiv V(\vec{r})/|V(\vec{r})|$  ( $\epsilon_V(\vec{r}) \equiv 0$  if  $V(\vec{r}) = 0$ ) and defining

$$\zeta_k^+(\vec{r}) \equiv |V(\vec{r})|^{\frac{1}{2}} \psi_k^+(\vec{r}) \quad (134)$$

$$K'(\vec{r}, \vec{r}') \equiv -\frac{2m}{\hbar^2} G_0(\vec{r}, \vec{r}') |V(\vec{r})|^{\frac{1}{2}} |V(\vec{r}')|^{\frac{1}{2}} \epsilon_V(\vec{r}'), \quad (135)$$

Equation 120 reads:

$$\zeta_k^+(\vec{r}) = |V(\vec{r})|^{\frac{1}{2}} \exp(i\vec{k} \cdot \vec{r}) + \lambda \int K'(\vec{r}, \vec{r}') \zeta_k^+(\vec{r}') d\vec{r}'. \quad (136)$$

Now the function in front of the integral is square integrable, and the kernel  $K'(\vec{r}, \vec{r}')$  is square integrable too:

$$\begin{aligned} & \int d\vec{r} \int d\vec{r}' |K'(\vec{r}, \vec{r}')|^2 \\ &= \int d\vec{r} \int d\vec{r}' \frac{|V(\vec{r})| |V(\vec{r}')|}{|\vec{r} - \vec{r}'|^2} < \infty \end{aligned} \quad (137)$$

provided the potential  $V(\vec{r})$  obeys the additional condition

$$\int d\vec{r}' \frac{|V(\vec{r}')|}{|\vec{r} - \vec{r}'|^2} < \infty. \quad (138)$$

Equation 136 can be formally written as

$$\zeta_k^+ = \zeta_0 + \lambda \hat{K}' \zeta_k^+, \quad \zeta_0 \equiv |V(\vec{r})|^{\frac{1}{2}} \exp(i\vec{k} \cdot \vec{r}), \quad (139)$$

$\hat{K}'$  being the integral operator with kernel  $K'$  given in Eq. 135. The function  $\zeta_k^+$  is formally given as

$$\zeta_k^+ = (I - \lambda \hat{K}')^{-1} \zeta_0 \quad (140)$$

where the inverse operator  $(I - \lambda \hat{K}')^{-1}$  exists except for those values of  $\lambda$  (singular values) for which  $I - \lambda \hat{K}'$  has the eigenvalue 0.

Since by Eq. 137  $\hat{K}'$  is a compact operator, the singular values are isolated points which obey the inequality  $|\lambda| \geq \|\hat{K}'\|^{-1}$ , since the spectrum of an operator is contained in the closed disk of radius equal to the norm of the operator. Thus, when  $\|\lambda \hat{K}'\| < 1$ , the inverse operator  $(I - \lambda \hat{K}')^{-1}$  exists and is given by the Neumann series

$$(I - \lambda \hat{K}')^{-1} = I + \lambda \hat{K}' + \lambda^2 \hat{K}'^2 + \dots \equiv I + R_\lambda^{K'}, \quad (141)$$

which is clearly norm convergent. By the inequality

$$\begin{aligned} \|\lambda \hat{K}'\|^2 &\leq \lambda^2 \text{Tr}(\hat{K}'^\dagger \hat{K}') \\ &= \lambda^2 \int d\vec{r} \int d\vec{r}' |K'(\vec{r}, \vec{r}')|^2, \end{aligned} \quad (142)$$

we see that, if

$$\begin{aligned} & \lambda^2 \frac{m^2}{4\pi^2 \hbar^4} \int d\vec{r} \int d\vec{r}' \frac{|V(\vec{r})| |V(\vec{r}')|}{|\vec{r} - \vec{r}'|^2} \\ & < 1, \end{aligned} \quad (143)$$

the condition  $\|\lambda \hat{K}'\| < 1$  is satisfied and consequently the inverse operator  $(I - \lambda \hat{K}')^{-1}$  exists. In conclusion, a sufficient condition for the convergence of the expansion



$$\zeta_k^+ = \zeta_0 + \lambda \hat{K}' \zeta_0 + \lambda^2 \hat{K}'^2 \zeta_0 + \dots \quad (144)$$

in the  $L^2$  norm is that Eq. 143 holds (Scadron et al. 1964). Since  $\|\hat{K}'\|^4 \leq \text{Tr}(\hat{K}'^\dagger \hat{K}' \hat{K}'^\dagger \hat{K}')$ , by the Riemann-Lebesgue lemma it is possible to prove (Zemach and Klein 1958) that for any given  $\lambda$ , the condition  $\|\lambda \hat{K}'\| \leq 1$  is satisfied provided the energy  $\hbar^2 k^2/2m$  is sufficiently large.

The implications for the convergence of the expansion Eq. 128 of the scattering amplitude are immediate, once Eq. 126 is written in the form

$$f_k^-(\theta, \varphi) = -\lambda \frac{m}{2\pi\hbar^2} \left( \left| V(\vec{r}) \right|^{\frac{1}{2}} \varphi_f(\vec{r}) \epsilon_V(\vec{r}), \right. \\ \left. \left| V(\vec{r}) \right|^{\frac{1}{2}} \psi_k^+(\vec{r}) \right). \quad (145)$$

The Born series converges whenever the iterative solution of Eq. 139 converges, that is, if Eq. 137 is satisfied. As noted above, for any given  $\lambda$  this happens for sufficiently large energy. An additional useful result is that the Born approximation  $f_k^B(\theta, \varphi)$  (or the expansion  $f_k^{[n]}(\theta, \varphi)$  through any  $n$ ) converges to the exact scattering amplitude  $f_k^-(\theta, \varphi)$  when the energy grows to infinity. More precisely (Thirring 2002), if Eq. 133 holds, then

$$\left| f_k^-(\theta, \varphi) - f_k^{[n]}(\theta, \varphi) \right| \xrightarrow{k \rightarrow \infty} 0. \quad (146)$$

If  $\|\lambda \hat{K}'\| \geq 1$ , the Neumann series Eq. 141 does not converge and the perturbation approach is no longer viable. However, if  $\lambda$  is not a singular value Eq. 139 can be solved by reducing it to an integral equation with a kernel  $D$  of norm less than 1 plus a problem of linear algebra (Yosida 1991). In fact, for any positive value  $L$ , the operator  $\hat{K}'$  can be approximated by a finite rank operator  $F$ :

$$F\zeta = \sum_{i=1}^n \alpha_i(\vec{r}) \left( \beta_i(\vec{r}'), \zeta(\vec{r}') \right) \quad (147)$$

such that if  $D \equiv \hat{K}' - F$ ,  $\|D\| < 1/L$ . Equation 139 then reads

$$(I - \lambda D)\zeta_k^+ = \zeta_0 + \lambda F\zeta_k^+. \quad (148)$$

Since for  $|\lambda| < L$  we have  $\|\lambda D\| < 1$ ,  $\zeta_k^+$  can be written in terms of the appropriate Neumann series  $I + R_\lambda^D$  (see Eq. 141):

$$\zeta_k^+ = \zeta_0 + R_\lambda^D \zeta_0 + F\zeta_k^+ + R_\lambda^D F\zeta_k^+. \quad (149)$$

The unknown quantities  $(\beta_i, \zeta_k^+)$  which appear in the RHS of Eq. 149 are determined by solving the linear-algebraic problem obtained by left multiplying both sides of the equation by  $\beta_r$ ,  $1 \leq r \leq n$ . The values of  $\lambda$  in the range  $|\lambda| < L$  for which the algebraic problem is not soluble are the singular values of Eq. 139 in that range. Thus, Eq. 139 can be solved for any nonsingular value.

## Time-Dependent Perturbations

A rather different problem is presented by the case that a time-independent Hamiltonian  $H_0$ , for which the spectrum and the eigenfunctions are known, is perturbed by a time-dependent potential  $V(t)$ . This occurs, for example, when an atom or a molecule interacts with an external electromagnetic field. For the total Hamiltonian  $H = H_0 + \lambda V(t)$ , stationary states no longer exist, and the relevant question is how the perturbation affects the time evolution of the system. We assume that the state  $\psi$  is known at a given time, which can be chosen as  $t = 0$ , and search for  $\psi(t)$ . Obviously, any time  $t_0$  prior to the setting on of the perturbation  $\lambda V(t)$  could be chosen instead of  $t = 0$ .

The time-dependent Schrödinger equation reads

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi(t) = H_0\psi(t) + \lambda V(t)\psi(t). \quad (150)$$

At any  $t$ ,  $\psi(t)$  can be expanded in the basis of the eigenfunctions  $\varphi_n(t)$  of  $H_0$ :

$$\begin{aligned}
 H_0 \varphi_n(t) &= E_n \varphi_n(t) \\
 \varphi_n(t) &= \varphi_n(0) \exp(-iE_n t/\hbar) \equiv \zeta_n \exp(-iE_n t/\hbar).
 \end{aligned}
 \tag{151}$$

For the sake of simplicity, we treat  $H_0$  as if it only had discrete spectrum, but the presence of a continuous component of the spectrum does not create any problem.

We can write (Dirac 1926; Sakurai 1967)

$$\psi(t) = \sum_n a_n(t) \varphi_n(t). \tag{152}$$

Note that the basis vectors  $\varphi_n(t)$  are themselves time dependent (by the phase factor given in Eq. 151), whereas the vectors  $\zeta_n$  are time independent. The isolation of the contribution of  $H_0$  to the time evolution as a time-dependent factor allows a simpler system of equations for the unknown coefficients  $a_n(t)$ .

Substituting expansion Eq. 152 into Eq. 150, we find

$$\begin{aligned}
 i\hbar \sum_n \dot{a}_n(t) \varphi_n(t) + \sum_n a_n(t) E_n \varphi_n(t) &= \\
 \sum_n a_n(t) E_n \varphi_n(t) + \lambda \sum_n a_n(t) V(t) \varphi_n(t).
 \end{aligned}$$

By left multiplying by  $\varphi_k(t)$ , for the coefficients  $a_k(t)$ , we find the system of equations

$$\begin{aligned}
 i\hbar \dot{a}_k(t) &= \lambda \sum_n a_n(t) (\varphi_k(t), V(t) \varphi_n(t)) \\
 &\equiv \lambda \sum_n V_{kn}^I(t) a_n(t),
 \end{aligned}
 \tag{153}$$

where we have defined

$$V_{kn}^I(t) = (\varphi_k(t), V(t) \varphi_n(t)). \tag{154}$$

The matrix elements  $V_{kn}^I(t)$  are the matrix elements of an operator  $V^I(t)$  between the time-independent vectors  $\zeta_k, \zeta_n$ . Indeed, since

$$\begin{aligned}
 \varphi_k(t) &= \zeta_n \exp(-iE_n t/\hbar) \\
 &= \exp(-iH_0 t/\hbar) \zeta_n,
 \end{aligned}
 \tag{155}$$

Eq. 154 can be written as

$$\begin{aligned}
 V_{kn}^I(t) &= \exp(-iH_0 t/\hbar) \zeta_k, V(t) \exp(-iH_0 t/\hbar) \zeta_n \\
 &\equiv (\zeta_k, V^I(t) \zeta_n),
 \end{aligned}
 \tag{156}$$

where the operator  $V^I(t)$  is defined as follows:

$$V^I(t) \equiv \exp(iH_0 t/\hbar) V(t) \exp(-iH_0 t/\hbar). \tag{157}$$

System Eq. 153 must be supplemented with the appropriate initial conditions, which depend on the particular problem. The commonest application of time-dependent perturbation theory is the calculation of transition probabilities between eigenstates of  $H_0$ . Thus, we assume that at  $t = 0$  the system is in an eigenstate of the unperturbed Hamiltonian  $H_0$ , say the state  $\varphi_1$ . In this case  $a_1(0) = 1, a_n(0) = 0$  if  $n \neq 1$ .

In the spirit of perturbation theory, each  $a_n$  is expanded into powers of  $\lambda$ :

$$\begin{aligned}
 a_1(t) &= 1 + \sum_{r=1} \lambda^r a_1^{(r)}(t), \\
 a_n(t) &= \sum_{r=1} \lambda^r a_n^{(r)}(t), \\
 n &\neq 1,
 \end{aligned}
 \tag{158}$$

and terms of equal order are equated. For  $a_k^{(r)}$ , we find

$$i\hbar \dot{a}_k^{(r)} = \sum_n V_{kn}^I a_n^{(r-1)}, \quad r > 0. \tag{159}$$

By Eq. 158, for any  $r > 0, a_n^{(r)}(0) = 0$ . As a consequence, for  $r = 1$  we have

$$\begin{aligned}
 a_k^{(1)} &= \frac{-i}{\hbar} \int_0^t V_{k1}^I(t_1) dt_1 \\
 &= \frac{-i}{\hbar} \int_0^t (\zeta_k, V(t_1) \zeta_1) \exp(i\Delta E_{k1} t_1/\hbar) dt_1,
 \end{aligned}
 \tag{160}$$

where  $\Delta E_{k1} \equiv E_k - E_1$ . For  $r = 2$ , we find

$$i\hbar \dot{a}_k^{(2)} = \sum_n V_{kn}^I(t) a_n^{(1)}(t) = \frac{-i}{\hbar} \sum_n V_{kn}^I(t) \int_0^t V_{n1}^I(t_1) dt_1$$

whose solution is

$$\begin{aligned}
 a_k^{(2)}(t) &= \left(\frac{-i}{\hbar}\right)^2 \int_0^t dt_2 \int_0^{t_2} dt_1 \sum_n V_{kn}^I(t_2) V_{n1}^I(t_1) \\
 &= \left(\frac{-i}{\hbar}\right)^2 \int_0^t dt_2 \int_0^{t_2} dt_1 \sum_n (\zeta_k, V(t_2) \zeta_n) \\
 &\quad \exp(i\Delta E_{kn} t_2 / \hbar) \times (\zeta_n, V(t_1) \zeta_1) \exp(i\Delta E_{n1} t_1 / \hbar).
 \end{aligned} \tag{161}$$

It is clear how the calculation proceeds for higher values of  $r$ . The general expression is

$$\begin{aligned}
 a_k^{(r)}(t) &= \left(\frac{-i}{\hbar}\right)^r \int_0^t dt_r \int_0^{t_r} dt_{r-1} \cdots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \times \\
 &\quad \sum V_{kn_r}^I(t_r) V_{n_r n_{r-1}}^I(t_{r-1}) \cdots V_{n_3 n_2}^I(t_2) V_{n_2 1}^I(t_1).
 \end{aligned} \tag{162}$$

By the completeness of the vectors  $\zeta_n$ , the sums over the intermediate states can be substituted by the identity, and the expression of  $a_k^{(r)}$  is simplified into

$$\begin{aligned}
 a_k^{(r)}(t) &= \left(\frac{-i}{\hbar}\right)^r \int_0^t dt_r \int_0^{t_r} dt_{r-1} \cdots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \\
 &\quad (\zeta_k, V^I(t_r) V^I(t_{r-1}) \cdots V^I(t_2) V^I(t_1) \zeta_1).
 \end{aligned} \tag{163}$$

It is customary to write Eq. 163 in a different way. The  $r$ -dimensional cube  $0 \leq t_i \leq t$ ,  $1 \leq i \leq r$  can be split into  $r!$  subdomains

$$0 \leq t_{p_1} \leq t_{p_2} \cdots \leq t_{p_{r-1}} \leq t_{p_r} \leq t, \tag{164}$$

with  $\{p_1, p_2, \dots, p_{r-1}, p_r\}$  a permutation of  $\{1, 2, \dots, r-1, r\}$ . The time-ordered product of  $r$  (noncommuting) operators  $V^I(t_{p_1}), V^I(t_{p_2}), \dots, V^I(t_{p_r})$  is introduced according to the definition

$$\begin{aligned}
 T[V^I(t_{p_1}) \cdots V^I(t_{p_r})] &\equiv V^I(t_r) \cdots V^I(t_1), \\
 t_1 \leq t_2 \leq \cdots \leq t_r.
 \end{aligned} \tag{165}$$

If  $(-i\lambda/\hbar)^r (\zeta_k, T[V^I(t_{p_1}) \cdots V^I(t_{p_r})] \zeta_1)$  is integrated over the  $r$ -cube, then each of the  $r!$  subdomains defined by Eq. 164 yields the same contribution. As a consequence, Eq. 162 can be written as

$$\begin{aligned}
 a_k^{(r)}(t) &= \frac{1}{r!} \left(\frac{-i}{\hbar}\right)^r \int_0^t dt_r \int_0^{t_r} dt_{r-1} \cdots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \\
 &\quad (\zeta_k, T[V^I(t_r) V^I(t_{r-1}) \cdots V^I(t_2) V^I(t_1)] \zeta_1).
 \end{aligned} \tag{166}$$

The amplitudes  $a_k(t)$  can then be written as

$$a_k(t) = \left( \zeta_k, T \left[ \exp \left( \frac{-i\lambda}{\hbar} \right) \int_0^t V^I(t') dt' \right] \zeta_1 \right) \tag{167}$$

with obvious significance of the  $T$ -exponential: each monomial in the  $V^I$  operators which appear in the expansion of the exponential is to be time ordered according to the  $T$ -prescription. If the initial state is given at time  $t_0$ , the integral appearing in the  $T$ -exponential should start at  $t_0$ . We define

$$U^I(t, t_0) \equiv T \left[ \exp \left( \frac{-i\lambda}{\hbar} \right) \int_{t_0}^t V^I(t') dt' \right]. \tag{168}$$

The expansion of the  $T$ -exponential into monomials in the  $V^I$  operators is called the Dyson series (Dyson 1949a, b). It is extensively employed in perturbative quantum field theory.

From Eqs. 167 and 168, it is easy to derive an expression for the time evolution operator  $U(t, t_0)$  such that

$$U(t, t_0) \psi(t_0) = \psi(t). \tag{169}$$

Indeed,  $\psi(t)$  and  $\psi(t_0)$  can be expanded in the basis of the vectors  $\varphi_n(t)$  and  $\varphi_n(t_0)$ , respectively, as in Eq. 152. By the linearity of the Schrödinger equation, it suffices to determine  $(\varphi_k(t), U(t, t_0) \varphi_n(t_0))$ , which we already know to be  $(\zeta_k, U^I(t, t_0) \zeta_n)$ . We have

$$\begin{aligned}
 (\varphi_k(t), U(t, t_0) \varphi_n(t_0)) &= \exp(-iH_0 t / \hbar) \zeta_k, \\
 &\quad U(t, t_0) \exp(-iH_0 t_0 / \hbar) \zeta_n \\
 &= (\zeta_k, \exp(iH_0 t / \hbar) U(t, t_0) \exp(-iH_0 t_0 / \hbar) \zeta_n).
 \end{aligned}$$

As a consequence, we find the equation

$$U(t, t_0) = \exp(-iH_0 t/\hbar) T \left[ \exp \left( \frac{-i\lambda}{\hbar} \right) \int_{t_0}^t V^I(t') dt' \right] \exp(iH_0 t_0/\hbar), \quad (170)$$

which provides the perturbation expansion of the evolution operator  $U(t, t_0)$  in powers of  $\lambda$ .

It can be proved that if the operator function  $V(t)$  is strongly continuous and the operators  $V(t)$  are bounded, then the expansion which defines the  $T$ -exponential is norm convergent to a unitary operator, as expected (Reed and Simon 1975). The restriction to bounded operators  $V(t)$  does not detract from the range of applications of Eqs. 167 and 170, since time-dependent perturbation theory is almost exclusively used for treating interactions of a system with external fields, which generate bounded interactions.

## Future Directions

The long and honorable service of perturbation theory in every sector of quantum mechanics must be properly acknowledged. Its future is perhaps already in our past: the main achievement is its application to quantum field theory where, just to quote an example, the agreement between the measured value and the theoretical prediction of the electron magnetic moment anomaly to ten significant digits has no rivals.

Despite its successes, still perturbation theory is confronted with fundamental questions. In most of realistic problems, it is unknown whether the perturbation series is convergent or at least asymptotic. In nonrelativistic quantum mechanics, this does not represent a practical problem since only a limited number of terms can be calculated, but in quantum field theory, where higher-order terms are in principle calculable, this calls for dedicated investigations. There, in particular, conditions for recovering the exact amplitudes from the first terms of the series by such techniques as the Padé approximants or the self-similar approximants, and the estimate on the bound of the error, deserve further investigation.

Somewhat paradoxically, it can be said that the future of perturbation theory is in the non-perturbative results (analyticity domains, large coupling constant behavior, tunneling effect. . .) – an issue where much work has already been done – since they have proved to be complementary to the use of perturbation theory.

## Bibliography

- Baker GA, Graves-Morris P (1996) Padé approximants. Cambridge University Press, Cambridge
- Bargmann V (1964) Note on Wigner's theorem on symmetry operations. *J Math Phys* 5:862–868
- Bender CM, Wu TT (1969) Anharmonic oscillator. *Phys Rev* 184:1231–1260
- Bloch C (1958) Sur la théorie des perturbations des états liés. *Nucl Phys* 6:329–347
- Böhm A (1993) Quantum mechanics, foundations and applications. Springer, New York, pp 208–215
- Borel E (1899) Mémoires sur le séries divergentes. *Ann Sci École Norm Supér* 16:9–136
- Born M (1926) Quantenmechanik der Stossvorgänge. *Z Phys* 38:803–827
- Born M, Heisenberg W, Jordan P (1926) Zur Quantenmechanik, II. *Z Phys* 35:557–615
- Bracci L, Picasso LE (2012) A simple iterative method to write the terms of any order of perturbation theory in quantum mechanics. *Eur Phys J Plus* 127:119
- Brillouin L (1932) Perturbation problem and self consistent field. *J Phys Radium* 3:373–389
- Courant R, Hilbert D (1989) Methods of mathematical physics, vol I. Wiley, New York, pp 343–350
- Dirac PAM (1926) On the theory of quantum mechanics. *Proc R Soc A* 112:661–677
- Dyson FJ (1949a) The radiation theories of Tomonaga, Schwinger and Feynman. *Phys Rev* 75:486–502
- Dyson FJ (1949b) The S-matrix in quantum electrodynamics. *Phys Rev* 75:1736–1755
- Dyson FJ (1952) Divergence of perturbation theory in quantum electrodynamics. *Phys Rev* 85:631–632
- Epstein ST (1954) Note on perturbation theory. *Am J Phys* 22:613–614
- Epstein ST (1968) Uniqueness of the energy in perturbation theory. *Am J Phys* 36:165–166
- Feynman RP (1939) Forces in molecules. *Phys Rev* 56:340–343
- Fock VA (1935) Zur Theorie des Wasserstoffatoms. *Z Phys* 98:145–154
- Graffi S, Grecchi V, Simon B (1970) Borel summability: application to the anharmonic oscillator. *Phys Lett B* 32:631–634
- Grossman A (1961) Schrödinger scattering amplitude I. *J Math Phys* 3:710–713
- Hamermesh M (1989) Group theory and its application to physical problems. Dover, New York, pp 32–114

- Hannabuss K (1997) Introduction to quantum theory. Clarendon, Oxford, pp 131–136
- Hellmann H (1937) Einführung in die Quantenchemie. Deuticke, Leipzig
- Joachain J (1983) Quantum collision theory. North-Holland, Amsterdam
- Kato T (1949) On the convergence of the perturbation method I. Prog Theor Phys 4:514–523
- Kato T (1966) Perturbation theory for linear operators. Springer, New York
- Kramers HA (1957) Quantum mechanics. North-Holland, Amsterdam, pp 198–202
- Krieger JB (1968) Asymptotic properties of perturbation theory. J Math Phys 9:432–435
- Landau LD, Lifshitz EM (1960) Mechanics. Pergamon Press, Oxford, p 129
- Lippmann BA, Schwinger J (1950) Variational principles for scattering processes I. Phys Rev 79:469–480
- Loeffel J, Martin A, Wightman A, Simon B (1969) Padé approximants and the anharmonic oscillator. Phys Lett B30:656–658
- Padé H (1899) Sur la représentation approchée d'une fonction pour des fonctions rationnelles. Ann Sci École Norm Supér 9(Suppl 3):1–93
- Rayleigh JW (1894–1896) The theory of sound, vol I. Macmillan, London, pp 115–118
- Reed M, Simon B (1975) Methods of modern mathematical physics, vol II. Academic, New York, pp 282–283
- Reed M, Simon B (1978) Methods of modern mathematical physics, vol IV. Academic, New York, pp 10–44
- Rellich F (1937a) Störungstheorie der Spektralzerlegung I. Math Ann 113:600–619
- Rellich F (1937b) Störungstheorie der Spektralzerlegung II. Math Ann 113:677–685
- Rellich F (1939) Störungstheorie der Spektralzerlegung III. Math Ann 116:555–570
- Rellich F (1940) Störungstheorie der Spektralzerlegung IV. Math Ann 117:356–382
- Riesz F, Sz.-Nagy B (1968) Leçons d'analyse fonctionnelle. Gauthier-Villars, Paris, pp 143–188
- Rollnik H (1956) Streumaxima und gebundene Zustände. Z Phys 145:639–653
- Sakurai JJ (1967) Advanced quantum mechanics. Addison-Wesley, Reading, pp 39–40
- Scadron M, Weinberg S, Wright J (1964) Functional analysis and scattering theory. Phys Rev 135:B202–B207
- Schrödinger E (1926) Quantisierung als Eigenwertproblem. Ann Phys 80:437–490
- Schwartz J (1960) Some non-self-adjoint operators. Commun Pure Appl Math 13:609–639
- Simon B (1970) Coupling constant analyticity for the anharmonic oscillator. Ann Phys 58:76–136
- Simon B (1991) Fifty years of eigenvalue perturbation theory. Bull Am Math Soc 24:303–319
- Thirring W (2002) Quantum mathematical physics. Springer, Berlin, p 177
- von Neumann J, Wigner E (1929) Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. Phys Z 30: 467–470
- Watson G (1912) A theory of asymptotic series. Philos Trans R Soc Lon Ser A 211:279–313
- Weyl H (1931) The theory of groups and quantum mechanics. Dover, New York
- Wigner EP (1935) On a modification of the Rayleigh-Schrödinger perturbation theory. Math Natur Anz (Budapest) 53:477–482
- Wigner EP (1959) Group theory and its application to the quantum mechanics of atomic spectra. Academic, New York
- Yosida K (1991) Lectures on differential and integral equations. Dover, New York, pp 115–131
- Yukalov VI, Yukalova EP (2007) Methods of self similar factor approximants. Phys Lett A 368:341–347
- Zemach C, Klein A (1958) The Born expansion in non-relativistic quantum theory I. Nuovo Cimento 10: 1078–1087

## Books and Reviews

- Hirschfelder JO, Byers Brown W, Epstein ST (1964) Recent developments in perturbation theory. In: Advances in quantum chemistry, vol 1. Academic, New York, pp 255–374
- Killingbeck J (1977) Quantum-mechanical perturbation theory. Rep Prog Phys 40:963–1031
- Mayer I (2003) Simple theorems, proofs and derivations in quantum chemistry. Kluwer Academic/Plenum Publishers, New York, pp 69–120
- Morse PM, Feshbach H (1953) Methods of theoretical physics, part 2. McGraw-Hill, New York, pp 1001–1106
- Wilcox CH (1966) Perturbation theory and its applications in quantum mechanics. Wiley, New York

## Normal Forms in Perturbation Theory

Henk W. Broer  
Instituut voor Wiskunde en Informatica,  
Rijksuniversiteit Groningen, Groningen, The  
Netherlands

### Article Outline

Glossary  
Definition of the Subject  
Introduction  
Motivation  
The Normal Form Procedure  
Background, Linearization  
Preliminaries from Differential Geometry  
Preservation of Structure  
Semi-local Normalization  
Applications  
Future Directions  
Bibliography

### Glossary

**Normal form procedure** This is the stepwise ‘simplification’ by changes of coordinates, of the Taylor series at an equilibrium point, or of similar series at periodic or quasi-periodic solutions.

**Preservation of structure** The normal form procedure is set up in such a way that all coordinate changes preserve a certain appropriate structure. This applies to the class of **Hamiltonian** or **volume preserving** systems, as well as to systems that are **equivariant** or **reversible** with respect to a symmetry group. In all cases the systems may also depend on **parameters**.

**Symmetry reduction** The truncated normal form often exhibits a toroidal symmetry that can be factored out, thereby leading to a lower dimensional reduction.

**Perturbation theory** The attempt to extend properties of the (possibly reduced) normal form truncation, to the full system.

### Definition of the Subject

Nonlinear dynamical systems are notoriously hard to tackle by analytic means. One of the few approaches that has been effective for the last couple of centuries, is Perturbation Theory. Here systems are studied, which in an appropriate sense, can be seen as perturbations of a given system with ‘well-known’ dynamical properties. Such ‘well-known’ systems usually are systems with a great amount of symmetry (like integrable Hamiltonian systems (Arnold 1980)) or very low-dimensional systems. The methods of Perturbation Theory then try to extend the ‘well-known’ dynamical properties to the perturbed system. Methods to do this are often based on the Implicit Function Theorem, on normal hyperbolicity (Takens and Vanderbauwhede 2009; Vanderbauwhede 1989) or on Kolmogorov–Arnold–Moser theory (Arnold 1980; Broer and Sevryuk 2008; Broer et al. 1990, 1996a; Ciocci et al. 2005).

To obtain a perturbation theory set-up, normal form theory is a vital tool. In its most elementary form it amounts to ‘simplifying’ the Taylor series of a dynamical system at an equilibrium point by successive changes of coordinates. The lower order truncation of the series often belongs to the class of ‘well-known’ systems and by the Taylor formula the original system then locally can be viewed as a perturbation of this ‘well-known’ truncation, that thus serves as the ‘unperturbed system’. More involved versions of normal form theory exist at periodic or quasi-periodic evolutions of the dynamical system.

### Introduction

We review the formal theory of normal forms of dynamical systems at equilibrium points. Systems

with continuous time, i.e., of vector fields, or autonomous systems of ordinary differential equations, are considered extensively. The approach is universal in the sense that it applies to many cases where a structure is preserved. In that case also the normalizing transformations preserve this structure. In particular this includes the Hamiltonian and the volume preserving case as well as cases that are equivariant or reversible with respect to a symmetry group. In all situations the systems may depend on parameters. Related topics are being dealt with concerning a vector field at a periodic solution or a quasi-periodic invariant torus as well as the case of a diffeomorphism at a fixed point. The paper is concluded by discussing a few non-formal aspects and some applications.

## Motivation

The term ‘normal form’ is widely used in mathematics and its meaning is very sensitive for the context. In the case of linear maps from a given vector space to itself, for example, one may consider all possible choices of a basis. Each choice gives a matrix-representation of a given linear map. A suitable choice of basis now gives the well-known Jordan canonical form. This normal form, in a simple way displays certain important properties of the linear map, concerning its spectrum, its eigenspaces, and so on.

Presently the concern is with dynamical systems, such as vector fields (i.e., systems of autonomous ordinary differential equations), or diffeomorphisms. The aim is to simplify these systems near certain equilibria and (quasi-periodic) solutions by a proper choice of coordinates. The purpose of this is to better display the local dynamical properties. This normalization is effected by a stepwise simplification of formal power series (such as Taylor series).

## Reduction of Toroidal Symmetry

In a large class of cases, the simplification of the Taylor series induces a toroidal symmetry up to a certain order, and truncation of the series at that order gives a local approximation of the system at hand. From this truncation we can reduce this

toroidal symmetry, thereby also reducing the dimension of the phase space. This kind of procedure is reminiscent of the classical reductions in classical mechanics related to the Noether Theorem, compare with (Arnold 1980), ► [“Dynamics of Hamiltonian Systems”](#).

This leads us to a first perturbation problem to be considered. Indeed, by truncating and factoring out the torus symmetry we get a polynomial system on a reduced phase, and one problem is how persistent this system is with respect to the addition of higher order terms. In the case where the system depends on parameters, the persistence of a corresponding bifurcation set is of interest.

In many examples the reduced phase space is 2-dimensional where the dynamics is qualitatively determined by a polynomial and this perturbation problem can be handled by Singularity Theory. In quite a number of cases the truncated and reduced system turns out to be structurally stable (Palis and de Melo 1982; Thom 1989).

## A Global Perturbation Theory

When returning to the original phase space we consider the original system as a perturbation of the de-reduced truncation obtained so far. In the ensuing perturbation problem, several types of resonance can play a role.

A classical example (Arnold 1983; Ciocci et al. 2005) occurs when in the reduced model a Hopf bifurcation takes place and we have reduced by a 1-torus. Then in the original space generically a Hopf–Neimark–Sacker bifurcation occurs, where in the parameter space the dynamics is organized by resonance *tongues*. In cases where we have reduced by a torus of dimension larger than 1, we are dealing with quasi-periodic Hopf bifurcation, in which the bifurcation set gets ‘Cantorized’ by Diophantine conditions, compare with (Braaksma and Broer 1987; Broer et al. 1990, 1998b; Ciocci et al. 2005). Also the theory of homo- and heteroclinic bifurcations then is of importance (Palis and Takens 1993).

We use the two perturbation problems as a motivation for the Normal Form Theory to be reviewed, for details, however, we just refer to the literature. For example, compare with ► [“Perturbation Theory”](#) and references therein.



## The Normal Form Procedure

The subject of our interest is the simplification of the Taylor series of a vector field at a certain equilibrium point. Before we explore this, however, let us first give a convenient normal form of a vector field near a non-equilibrium point.

**Theorem 1 (Flow box (Spivak 1970))** *Let the  $C^\infty$  vector field  $X$  on  $\mathbb{R}^n$  be given by  $\dot{x} = f(x)$  and assume that  $f(p) \neq 0$ . Then there exists a neighborhood of  $p$  with local  $C^\infty$  coordinates  $y = (y_1, y_2, \dots, y_n)$ , such that in these coordinates  $X$  has the form*

$$\begin{aligned}\dot{y}_1 &= 1 \\ \dot{y}_j &= 0,\end{aligned}$$

for  $2 \leq j \leq n$ .

Such a local chart usually is called a *flowbox* and the above theorem the Flowbox Theorem. A proof simply can be given using a transversal local  $C^\infty$  section that cuts the flow of the vector field transversally. For the coordinate  $y_1$  then use the time-parametrization of the flow, while the coordinates  $y_j$ , for  $2 \leq j \leq n$  come from the section, compare with, e.g., (Spivak 1970).

## Background, Linearization

The idea of simplification near an equilibrium goes back at least to Poincaré (1928), also compare with Arnold (1983). To be definite, we now let  $X$  be a  $C^\infty$  vector field on  $\mathbb{R}^n$ , with the origin as an equilibrium point. Suppose that  $X$  has the form  $\dot{x} = Ax + f(x)$ ,  $x \in \mathbb{R}^n$ , where  $A$  is linear and where  $f(0) = 0$ ,  $D_0 f = 0$ . The first idea is to apply successive  $C^\infty$  changes of coordinates of the form  $\text{Id} + P$ , with  $P$  a homogeneous polynomial of degree  $m = 2, 3, \dots$ , ‘simplifying’ the Taylor series step by step.

The most ‘simple’ form that can be obtained in this way, is where *all* higher order terms vanish. In that case the normal form is formally linear. Such a case was treated by Poincaré, and we shall investigate this now.

We may even assume to work on  $\mathbb{C}^n$ , for simplicity assuming that the eigenvalues of  $A$  are

distinct. A collection  $\lambda = (\lambda_1, \dots, \lambda_n)$  of points in  $\mathbb{C}$  is said to be *resonant* if there exists a relation of the form

$$\lambda_s = \langle r, \lambda \rangle,$$

for  $r = (r_1, \dots, r_n) \in \mathbb{Z}^n$ , with  $r_k \geq 0$  for all  $k$  and with  $\sum r_k \geq 2$ . The *order* of the resonance then is the number  $|r| = \sum r_k$ . The Poincaré Theorem now reads

**Theorem 2 (Formal linearization (Arnold 1983; Poincaré 1928))** *If the (distinct) eigenvalues  $\lambda_1, \dots, \lambda_n$  of  $A$  have no resonances, there exists a formal change of variables  $x = y + O(|y|^2)$ , transforming the above vector field  $X$ , given by*

$$\dot{x} = Ax + f(x)$$

to

$$\dot{y} = Ay.$$

We include a proof (Arnold 1983), since this will provide the basis for almost all further considerations.

**Proof** The formal power series  $x = y + O(|y|^2)$  is obtained in an inductive manner. Indeed, for  $m = 2, 3, \dots$  a polynomial transformation  $x = y + P(y)$  is constructed, with  $P$  homogeneous of degree  $m$ , which removes the terms of degree  $m$  from the vector field. At the end we have to take the composition of all these polynomial transformations.

1. The basic tool for the  $m$ th step is the following. Let  $v$  be homogeneous of degree  $m$ , then if the vector fields  $\dot{x} = Ax + v(x) + O(|x|^{m+1})$  and  $\dot{y} = Ay$  are related by the transformation  $x = y + P(y)$  with  $P$  also homogeneous of degree  $m$ , then

$$D_x P A x - A P(x) = v(x).$$

This relation usually is called the *homological equation*, the idea being to determine  $P$  in terms of

$v$ : by this choice of  $P$  the term  $v$  can be transformed away.

The proof of this relation is straightforward. In fact,

$$\begin{aligned}\dot{x} &= (\text{Id} + D_y P) A y \\ &= (\text{Id} + D_y P) A (x - P(x) + O(|x|^{m+1})) \\ &= Ax + \{D_x P A x - A P(x)\} + O(|x|^{m+1}),\end{aligned}$$

where we used that for the inverse transformation we know  $y = x - P(x) + O(|x|^{m+1})$ .

2. For notational convenience we introduce the linear operator  $\text{ad}A$ , the so-called adjoint operator, by

$$\text{ad}A(P)(x) := D_x P A x - A P(x),$$

then the homological equation reads  $\text{ad}A(P) = v$ . So the question is reduced to whether  $v$  is in the image of the operator  $\text{ad}A$ . It turns out that the eigenvalues of  $\text{ad}A$  can be expressed in those of  $A$ . If  $x_1, x_2, \dots, x_n$  are the coordinates corresponding to the basis  $e_1, e_2, \dots, e_n$ , again it is a straightforward computation to show that for  $P(x) = x^r e_s$ , one has

$$\text{ad}A(P) = (\langle r, \lambda \rangle - \lambda_s) P.$$

Here we use the multi-index notation  $x^r = x_1^{r_1} x_2^{r_2} \dots x_n^{r_n}$ . Indeed, for this choice of  $P$  one has  $AP(x) = \lambda_s P(x)$ , while

$$\frac{\partial x^r}{\partial x} A x = \sum_j \frac{r_j}{x_j} x^r \lambda_j x_j = \langle r, \lambda \rangle x^r.$$

We conclude that the monomials  $x^r e_s$  are eigenvectors corresponding to the eigenvalues  $\langle r, \lambda \rangle - \lambda_s$ . We conclude that the operator  $\text{ad}A$  is *semi-simple*, since it has a basis of eigenvectors. Therefore, if  $\ker \text{ad}A = 0$ , the operator is surjective. This is exactly what the non-resonance condition on the eigenvalues of  $A$  amounts to. More precisely, the homological equation  $\text{ad}A(P) = v$  can be solved for  $P$  for each homogeneous part  $v$  of

degree  $m$ , provided that there are no resonances up to order  $m$ .

3. The induction process now runs as follows. For  $m \geq 2$ , given a form

$$\dot{x} = Ax + v_m(x) + O(|x|^{m+1}),$$

we solve the homological equation

$$\text{ad}A(P_m) = v_m,$$

then carrying out the transformation  $x = y + P_m(y)$ . This takes the above form to

$$\dot{y} = Ay + w_{m+1}(y) + O(|y|^{m+2}).$$

The composition of all the polynomial transformations then gives the desired formal transformation.

#### Remark

- It is well-known that the formal series usually diverge. Here we do not go into this problem, for a brief discussion see below.
- If resonances are excluded up to a finite order  $N$ , we can linearize up to that order, so obtaining a normal form.

$$\dot{y} = Ay + O(|y|^{N+1}).$$

In this case the transformation can be taken as a polynomial.

- If the original problem is real, but with the matrix  $A$  having non-real eigenvalues, we still can keep all transformations real by also considering complex conjugate eigenvectors.

We conclude this introduction by discussing two further linearization theorems, one due to Sternberg and the other to Hartman–Grobman. We recall the following for a vector field  $X(x) = Ax + f(x)$ ,  $x \in \mathbb{R}^n$ , with 0 as an equilibrium point, i.e., with  $f(0) = 0$ ,  $D_0 f = 0$ . The equilibrium 0 is

*hyperbolic* if the matrix  $A$  has no purely imaginary eigenvalues. Sternberg's Theorem reads

**Theorem 3 (Smooth linearization (Sternberg 1959))** *Let  $X$  and  $Y$  be  $C^\infty$  vector fields on  $\mathbb{R}^n$ , with  $0$  as a hyperbolic equilibrium point. Also suppose that there exists a formal transformation  $(\mathbb{R}^n, 0) \rightarrow (\mathbb{R}^n, 0)$  taking the Taylor series of  $X$  at  $0$  to that of  $Y$ . Then there exists a local  $C^\infty$ -diffeomorphism  $\Phi : (\mathbb{R}^n, 0) \rightarrow (\mathbb{R}^n, 0)$ , such that  $\Phi_*X = Y$ .*

We recall that  $\Phi_*X(\Phi(x)) = D_x\Phi X(x)$ . This means that  $X$  and  $Y$  are locally *conjugated* by  $\Phi$ , the evolution curves of  $X$  are mapped to those of  $Y$  in a time-preserving manner. In particular Sternberg's Theorem applies when the conclusion of Poincaré's Theorem holds: for  $Y$  just take the linear part  $Y(x) = Ax(\partial/\partial x)$ .

Combining these two theorems we find that in the hyperbolic case, under the exclusion of all resonances, the vector field  $X$  is linearizable by a  $C^\infty$ -transformation. The Hartman–Grobman Theorem moreover says that the non-resonance condition can be omitted, provided we only want a  $C^0$ -linearization.

**Theorem 4 (Continuous linearization e.g., (Palis and de Melo 1982))** *Let  $X$  be a  $C^\infty$  vector field on  $\mathbb{R}^n$ , with  $0$  as a hyperbolic equilibrium point. Then there exists a local homeomorphism  $\Phi : (\mathbb{R}^n, 0) \rightarrow (\mathbb{R}^n, 0)$ , locally conjugating  $X$  to its linear part.*

## Preliminaries from Differential Geometry

Before we develop a more general Normal Form Theory we recall some elements from differential geometry. One central notion used here is that of the *Lie derivative*. For simplicity all our objects will be of class  $C^\infty$ . Given a vector field  $X$  we can take any tensor  $\mathbf{t}$  and define its Lie-derivative  $\mathcal{L}_X\mathbf{t}$  with respect to  $X$  as the infinitesimal transformation of  $\mathbf{t}$  along the flow of  $X$ . In this way  $\mathcal{L}_X\mathbf{t}$  becomes a tensor of the same type as  $\mathbf{t}$ . To be more precise, for  $\tau$  a real *function* one so defines

$$\mathcal{L}_X f(x) = X(f)(x) = \left. \frac{d}{dt} f(X_t(x)) \right|_{t=0},$$

i.e., the directional derivative of  $f$  with respect to  $X$ . Here  $X_t$  denotes the flow of  $X$  over time  $t$ . For  $\tau$  a vector field  $Y$  one similarly defines

$$\begin{aligned} \mathcal{L}_X Y(x) &= \left. \frac{d}{dt} (X_{-t})_* Y(x) \right|_{t=0} \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \{ (X_{-t})_* Y(x) - Y(x) \} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \{ Y(x) - (X_h)_* Y(x) \}, \end{aligned}$$

and similarly for *differential forms*, etc. Another central notion is the *Lie-brackets*  $[X, Y]$  defined for any two vector fields  $X$  and  $Y$  on  $\mathbb{R}^n$  by

$$[X, Y](f) = X(Y(f)) - Y(X(f)).$$

Here  $f$  is any real function on  $\mathbb{R}^n$  while, as before,  $X(f)$  denotes the directional derivative of  $f$  with respect to  $X$ .

We recall the expression of Lie-brackets in coordinates. If  $X$  is given by the system of differential equations  $\dot{x}_j = X_j(x)$ , with  $1 \leq j \leq n$ , then the directional derivative  $X(f)$  is given by  $X(f) = \sum_{j=1}^n X_j \partial f / \partial x_j$ . Then, if  $[X, Y]$  is given by the system  $\dot{x}_j = Z_j(x)$ , for  $1 \leq j \leq n$ , of differential equations, one directly shows that

$$Z_j = \sum_{k=1}^n \left( X_k \frac{\partial Y_j}{\partial x_k} - Y_k \frac{\partial X_j}{\partial x_k} \right).$$

Here  $Y_j$  relates to  $Y$  as  $X_j$  does to  $X$ . We list some useful properties.

**Proposition 5 (Properties of the Lie-derivative (Spivak 1970))**

1.  $\mathcal{L}_X(Y_1 + Y_2) = \mathcal{L}_X Y_1 + \mathcal{L}_X Y_2$  (*linearity over  $R$* )
2.  $\mathcal{L}_X(fY) = X(f) \times Y + f \times \mathcal{L}_X Y$  (*Leibniz rule*)
3.  $[Y, X] = -[X, Y]$  (*skew symmetry*)
4.  $[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0$  (*Jacobi identity*)
5.  $\mathcal{L}_X Y = [X, Y]$
6.  $[X, Y] = 0 \Leftrightarrow X_t \circ Y_s = Y_s \circ X_t$

**Proof** The first four items are left to the reader.

5. The equality  $\mathcal{L}_X Y = [X, Y]$  can be proven by observing the following. Both members of the equality are defined intrinsically, so it is enough to check it in any choice of (local) coordinates. Moreover, we can restrict our attention to the set  $\{x \mid X(x) \neq 0\}$ . By the Flowbox Theorem 1 we then may assume for the component functions of  $X$  that  $X_1(x) = 1$ ,  $X_j(x) = 0$  for  $2 \leq j \leq n$ . It is easy to see that both members of the equality now are equal to the vector field  $Z$  with components

$$Z_j = \frac{\partial Y_j}{\partial x_1}.$$

6. Remains the equivalence of the commuting relationships. The commuting of the flows, by a very general argument, implies that the bracket vanishes. In fact, fixing  $t$  we see that  $X_t$  conjugates the flow of  $Y$  to itself, which is equivalent to  $(X_t)_* Y = Y$ . By definition this implies that  $\mathcal{L}_X Y = 0$ .

Conversely, let  $c(t) = ((X_t)_* Y)(p)$ . From the fact that  $\mathcal{L}_X Y = 0$  it then follows that  $c(t) \equiv c(0)$ . Observe that the latter assertion is sufficient for our purposes, since it implies that  $(X_t)_* Y = Y$  and therefore  $X_t \circ Y_s = Y_s \circ X_t$ . Finally, that  $c(t)$  is constant can be shown as follows.

$$\begin{aligned} c'(t) &= \lim_{h \rightarrow 0} \frac{1}{h} \{c(t+h) - c(t)\} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \{((X_{t+h})_* Y)(p) - ((X_t)_* Y)(p)\} \\ &= (X_t)_* \lim_{h \rightarrow 0} \frac{1}{h} \{((X_h)_* Y)(X_{-t}(p)) - Y(X_{-t}(p))\} \\ &= (X_t)_* \mathcal{L}_X Y(X_{-t}(p)) \\ &= (X_t)_*(0) \\ &= 0. \end{aligned}$$

### ‘Simple’ in Terms of an Adjoint Action

We now return to the setting of Normal Form Theory at equilibrium points. So given is the vector field  $\dot{x} = X(x)$ , with  $X(x) = Ax + f(x)$ ,  $x \in \mathbb{R}^n$ , where  $A$  is linear and where  $f(0) = 0$ ,  $D_0 f = 0$ . We recall that it is our general aim to ‘simplify’ the Taylor series of  $X$  at 0.

However, we have not yet said what the word ‘simple’ means in the present setting. In order to understand what is going on, we reintroduce the adjoint action associated to the linear part  $A$ ,

defined on the class of all  $C^\infty$  vector fields on  $\mathbb{R}^n$ . To be precise, this adjoint action  $\text{ad}A$  is defined by the Lie-bracket

$$\text{ad}A : Y \mapsto [A, Y],$$

where  $A$  is identified with the linear vector field  $\dot{x} = Ax$ . It is easily seen that this fits with the notation introduced in Theorem 2.

Let  $H^m(\mathbb{R}^n)$  denote the space of polynomial vector fields, homogeneous of degree  $m$ . Then the Taylor series of  $X$  can be viewed as an element of the product  $\prod_{m=1}^{\infty} H^m(\mathbb{R}^n)$ . Also, it directly follows that  $\text{ad}A$  induces a linear map  $H^m(\mathbb{R}^n) \rightarrow H^m(\mathbb{R}^n)$ , to be denoted by  $\text{ad}_m A$ . Let

$$B^m := \text{im ad}_m A,$$

the image of the map  $\text{ad}_m A$  in  $H^m(\mathbb{R}^n)$ . Then for any complement  $G^m$  of  $B^m$  in  $H^m(\mathbb{R}^n)$ , in the sense that

$$B^m \oplus G^m = H^m(\mathbb{R}^n),$$

we define the *corresponding* notion of ‘simplicity’ by requiring the homogeneous part of degree  $m$  to be in  $G^m$ . In the case of the Poincaré Theorem 2, since  $B^m = H^m(\mathbb{R}^n)$ , we have  $G^m = \{0\}$ .

We now quote a theorem from Sect. 7.6.1. in Dumortier et al. (Broer et al. 1991). Although its proof is very similar to the one of Theorem 2, we include it here, also because of its format.

**Theorem 6 (‘Simple’ in terms of  $G^m$  (Roussarie 1987; Sternberg 1959))** *Let  $X$  be a  $C^\infty$  vector field, defined in the neighborhood of  $0 \in \mathbb{R}^n$ , with  $X(0) = 0$  and  $D_0 X = A$ . Also let  $N \in \mathbb{N}$  be given and, for  $m \in \mathbb{N}$ , let  $B^m$  and  $G^m$  be such that  $B^m \oplus G^m = H^m(\mathbb{R}^n)$ . Then there exists, near  $0 \in \mathbb{R}^n$ , an analytic change of coordinates  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , with  $\Phi(0) = 0$ , such that*

$$\begin{aligned} \Phi_* X(y) &= Ay + g_2(y) + \cdots + g_N(y) \\ &\quad + O(|y|^{N+1}), \end{aligned}$$

with  $g_m \in G^m$ , for all  $m = 2, 3, \dots, N$ .

**Proof** We use induction on  $N$ . Let us assume that

$$X(x) = Ax + g_2(x) + \cdots + g_{N-1}(x) + f_N(x) + O(|x|^{N+1}),$$

with  $g_m \in G^m$ , for all  $m = 2, 3, \dots, N-1$  and with  $f_N$  homogeneous of degree  $N$ .

We consider a coordinate change  $x = y + P(y)$ , where  $P$  is polynomial of degree  $N$ , see above. For any such  $P$ , by substitution we get

$$(\text{Id} + D_y P)\dot{y} = A(y + P(y)) + g_2(y) + \cdots + g_{N-1}(y) + f_N(y) + O(|y|^{N+1}),$$

or

$$\begin{aligned} \dot{y} &= (\text{Id} + D_y P)^{-1} \left( A(y + P(y)) + g_2(y) + \cdots + g_{N-1}(y) + f_N(y) + O(|y|^{N+1}) \right) \\ &= Ay + g_2(y) + \cdots + g_{N-1}(y) + f_N(y) + AP(y) \\ &\quad - D_y P Ay + O(|y|^{N+1}), \end{aligned}$$

using that  $(\text{Id} + D_y P)^{-1} = \text{Id} - D_y P + O(|y|^N)$ . We conclude that the terms up to order  $N-1$  are unchanged by this transformation, while the  $N$ th order term becomes

$$f_N(y) - \text{ad}_N A(P)(y).$$

Clearly, a suitable choice of  $P$  will put this term in  $G^N$ . This is the present version of the *homological equation* as introduced before.

**Remark** For simplicity the formulations all are in the  $C^\infty$ -context, but obvious changes can be made for the case of finite differentiability. The latter case is of importance for applications of Normal Form Theory after reduction to a center manifold (Takens and Vanderbauwhede 2009; Vanderbauwhede 1989).

### Torus Symmetry

As a special case let the linear part  $A$  be semisimple, in the sense that it is diagonalizable over the complex numbers. It then directly follows that also  $\text{ad}_m A$  is semisimple, which implies that

$$\text{im ad}_m A \oplus \ker \text{ad}_m A = H^m(\mathbb{R}^n).$$

The reader is invited to provide the eigenvalues of  $\text{ad}_m A$  in terms of those of  $A$ , compare with the proof of Theorem 2. In the present case the obvious choice for the complementary space defining

‘simpleness’ is  $G^m = \ker \text{ad}_m A$ . Moreover, the fact that the normalized, viz. simplified, terms  $g_m$  are in  $G^m$  by definition means that

$$[A, g_m] = 0.$$

This, in turn, implies that  $N$ -jet of  $\Phi_* X$ , i.e., the *normalized part* of  $\Phi_* X$ , by Proposition 5 is invariant under all linear transformations

$$\exp tA, \quad t \in \mathbb{R}$$

generated by  $A$ . For further reading also compare with Sternberg (1959), Takens (1974), Broer (1981, 1993) or (Broer and Takens 1989).

More generally, let  $A = A_s + A_n$  be the Jordan canonical splitting in the semisimple and nilpotent part. Then one directly shows that  $\text{ad}_m A = \text{ad}_m A_s + \text{ad}_m A_n$  is the Jordan canonical splitting, whence, by a general argument (Varadarajan 1974) it follows that

$$\text{im ad}_m A + \ker \text{ad}_m A_s = H^m(\mathbb{R}^n),$$

so where the sum splitting in general no longer is direct. Now we can choose the complementary spaces  $G^m$  such that

$$G^m \subset \ker \text{ad}_m A_s,$$

ensuring equivariance of the normalized part of  $\Phi_* X$ , with respect to all linear transformations

$$\exp tA_s, \quad t \in \mathbb{R}.$$

The choice of  $G^m$  can be further restricted, e.g., such that

$$G^m \subseteq \ker \operatorname{ad}_m A_s \setminus \operatorname{im} \operatorname{ad}_m A_n \subseteq H^m(\mathbb{R}^n),$$

compare with Van der Meer (1985). For further discussion on the choice of  $G^m$ , see below.

**Example (Rotational symmetry (Takens 1974))**

Consider the case  $n = 2$  where

$$A = A_s = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

So, the eigenvalues of  $A$  are  $\pm i$  and the transformations  $\exp t A_s$ ,  $t \in \mathbb{R}$ , form the rotation group  $\operatorname{SO}(2, \mathbb{R})$ . From this, we can arrive at once at the general format of the normal form. In fact, the normalized,  $N$ th order part of  $\Phi_* X$  is rotationally symmetric. This implies that, if we pass to polar coordinates  $(r, \varphi)$ , by

$$y_1 = r \cos \varphi, \quad y_2 = r \sin \varphi,$$

the normalized truncation of  $\Phi_* X$  obtains the form

$$\begin{aligned} \dot{\varphi} &= f(r^2) \\ \dot{r} &= rg(r^2), \end{aligned}$$

for certain polynomials  $f$  and  $g$ , with  $f(0) = 1$  and  $g(0) = 0$ .

**Remark**

- A more direct, ‘computational’ proof of this result can be given as follows, compare with the proof of Theorem 2 and with (Broer et al. 1991; Takens 1974; Vanderbauwhede 1989). Indeed, in vector field notation we can write

$$\begin{aligned} A &= -x_2 \frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial x_2} \\ &= i \left( z \frac{\partial}{\partial z} - \bar{z} \frac{\partial}{\partial \bar{z}} \right), \end{aligned}$$

where we complexified putting  $z = x_1 + ix_2$  and use the well-known Wirtinger derivatives

$$\frac{\partial}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x_1} - i \frac{\partial}{\partial x_2} \right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left( \frac{\partial}{\partial x_1} + i \frac{\partial}{\partial x_2} \right).$$

Now a basis of eigenvectors for  $\operatorname{ad}_m A$  can be found directly, just computing a few Lie-brackets, compare the previous subsection. In fact, it is now given by all monomials

$$z^k \bar{z}^\ell \frac{\partial}{\partial z}, \quad \text{and} \quad z^k \bar{z}^\ell \frac{\partial}{\partial \bar{z}},$$

with  $k + \ell = m$ . The corresponding eigenvalues are  $i(k - \ell - 1)$  viz.  $i(k - \ell + 1)$ . So again we see, now by a direct inspection, that  $\operatorname{ad}_m A$  is semi-simple and that we can take  $G^m = \ker \operatorname{ad}_m A$ . This space is spanned by

$$(z\bar{z})^r \left( z \frac{\partial}{\partial z} \pm \bar{z} \frac{\partial}{\partial \bar{z}} \right),$$

with  $2r + 1 = m - 1$ , which indeed proves that the normal form is rotationally symmetric.

- A completely similar case occurs for  $n = 3$ , where

$$A = A_s = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In this case the normalized part in cylindrical coordinates  $(r, \varphi, z)$ , given by

$$y_1 = r \cos \varphi, \quad y_2 = r \sin \varphi, \quad y_3 = z,$$

in general gets the axially symmetric form

$$\begin{aligned} \dot{\varphi} &= f(r^2, z) \\ \dot{r} &= rg(r^2, z) \\ \dot{z} &= h(r^2, z), \end{aligned}$$

for suitable polynomials  $f$ ,  $g$  and  $h$ .

As a generalization of this example we state the following proposition, where the normalized part exhibits an  $m$ -torus symmetry.

**Proposition 7 (Toroidal symmetry (Takens 1974))** *Let  $X$  be a  $C^\infty$  vector field, defined in the*

neighborhood of  $0 \in \mathbb{R}^n$ , with  $X(0) = 0$  and where  $A = D_0X$  is semisimple with the eigenvalues  $\pm i\omega_1, \pm i\omega_2, \dots, \pm i\omega_m$  and  $0$ . Here  $2m \leq n$ . Suppose that for given  $N \in \mathbb{N}$  and all integer vectors  $(k_1, k_2, \dots, k_m)$ ,

$$1 \leq \sum_{j=1}^m |k_j| \leq N+1 \Rightarrow \sum_{j=1}^m k_j \omega_j \neq 0, \quad (1)$$

(i.e., there are no resonances up to order  $N+1$ ). Then there exists, near  $0 \in \mathbb{R}^n$ , an analytic change of coordinates  $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , with  $\Phi(0) = 0$ , such that  $\Phi_*X$ , up to terms of order  $N$  has the following form. In suitable generalized cylindrical coordinates  $(\varphi_1, \dots, \varphi_m, r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n)$  it is given by

$$\begin{aligned} \dot{\varphi}_j &= f_j(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n) \\ \dot{r}_j &= r_j g_j(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n) \\ \dot{z}_\ell &= h_\ell(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n), \end{aligned}$$

where  $f_j(0) = \omega_j$  and  $h_\ell(0) = 0$  for  $1 \leq j \leq m, n-2m+1 \leq \ell \leq n$ .

Proofs can be found, e.g., in (Broer et al. 1991; Takens 1974; Takens and Vanderbauwhede 2009; Vanderbauwhede 1989). In fact, if one introduces a suitable complexification, it runs along the same lines as the above remark. For the fact that finitely many non-resonance conditions are needed in order to normalize up to finite order, also compare a remark following Theorem 2.

Since the truncated system of  $\dot{r}_j$ - and  $\dot{z}_\ell$ -equations is independent of the angles  $\varphi_j$ , this can be studied separately. A similar remark holds for the earlier examples of this section. As indicated in the introduction, this kind of ‘reduction by symmetry’ to lower dimension can be of great importance when studying the dynamics of  $X$ : as it enables us to consider  $X$ , viz.  $\Phi_*X$ , as an  $N$ -flat perturbation of the normalized part, which is largely determined by this lower dimensional reduction. For more details see below.

### On the Choices of the Complementary Space and of the Normalizing Transformation

In the previous subsection we only provided a general (symmetric) format of the normalized part. In concrete examples one has to do more.

Indeed, given the original Taylor series, one has to compute the coefficients in the normalized expansion. This means that many choices have to be made explicit.

To begin with there is the choice of the spaces  $G^m$ , which define the notion of ‘simple’. We have seen already that this choice is not unique.

Moreover observe that, even if the choice of  $G^m$  has been fixed, still  $P$  usually is not uniquely determined. In the semisimple case, for example,  $P$  is only determined modulo the kernel  $G^m = \ker \operatorname{ad}_m A$ .

**Remark** To fix thoughts, consider the former of the above examples, on  $\mathbb{R}^2$ , where the normalized truncation has the rotationally symmetric form

$$\begin{aligned} \dot{\varphi} &= f(r^2) \\ \dot{r} &= rg(r^2). \end{aligned}$$

Here  $g(r^2) = cr^2 + O(|r|^4)$ . The coefficient  $c$  dynamically is important, just think of the case where the system is part of a family that goes through Hopf-bifurcation. The computation of (the sign of)  $c$  in a concrete model can be quite involved, as it appears from, e.g., Marsden and McCracken (1976). Machine-assisted methods largely have taken over this kind of work.

One general way to choose  $G^m$  is the following, compare with e.g., Sect. 7.6 in Broer et al. (1991) and Sect. 2.3 in Vanderbauwhede (1989):

$$G^m := \ker(\operatorname{ad}_m A^T).$$

Here  $A^T$  is the *transpose* of  $A$ , defined by the relation  $\langle A^T x, y \rangle = \langle x, Ay \rangle$ , where  $\langle \cdot, \cdot \rangle$  is an inner product on  $\mathbb{R}^n$ . A suitable choice for an inner product on  $H^m(\mathbb{R}^n)$  then directly gives that

$$G^m \oplus \operatorname{im}(\operatorname{ad}_m A) = H^m(\mathbb{R}^n),$$

as required. Also here the normal form can be interpreted in terms of symmetry, namely with respect to the group generated by  $A^T$ . In the semisimple case, this choice leads to exactly the same symmetry considerations as before.

The above algorithms do not provide methods for computing the normal form yet, i.e., for



actually solving the homological equation. In practice, this is an additional computation. Regarding the corresponding algorithms we give a few more references for further reading, also referring to their bibliographies. General work in this direction is Bruno (1989, 2000), Sect. 7.6 in Dumortier et al. (Broer et al. 1991), Takens (1974) or Vanderbauwhede (1989), Part 2. In the latter reference also a brief description is given of the  $\mathfrak{sl}(2, \mathbb{R})$ -theory of Sanders and Cushman (Cushman and Sanders 1986), which is a powerful tool in the case where the matrix  $A$  is nilpotent (Takens 1974). For a thorough discussion on some of these methods we refer to Murdock (2003). Later on we shall come back to these aspects.

## Preservation of Structure

It goes without saying that Normal Form Theory is of great interest in special cases where a given structure has to be preserved. Here one may think of a symplectic or a volume form that has to be respected. Also a given symmetry group can have this role, e.g., think of an involution related to reversibility. Another, similar problem is the dependence of external parameters in the system.

A natural language for preservation of such structures is that of Lie-subalgebras of general Lie-algebra of vector fields, and the corresponding Lie-subgroup of the general Lie-group of diffeomorphisms.

### The Lie-Algebra Proof

Fortunately the setting of Theorem 6 is almost completely in terms of Lie-brackets. Let us briefly reconsider its proof.

Given is a  $C^\infty$  vector field  $X(x) = Ax + f(x)$ ,  $x \in \mathbb{R}^n$ , where  $A$  is linear and where  $f(0) = 0$ ,  $D_0 f = 0$ . We recall that in the inductive procedure a transformation

$$h = \text{Id} + P,$$

with  $P$  a homogeneous polynomial of degree  $m = 2, 3, \dots$ , is found, putting the homogeneous  $m$ th

degree part of the Taylor series of  $X$  into the ‘good’ space  $G^m$ .

Now, nothing changes in this proof if instead of  $h = \text{Id} + P$ , we take  $h = P_1$ , the flow over time 1 of the vector field  $P$ : indeed, the effect of this change is not felt until the order  $2m - 1$ . Here we use the following formula for  $X^t := (P_t)_* X$ :

$$[X^t, P] = \text{ad}_m A(P) + O(|y|^{m+1})$$

$$\text{and } \frac{\partial X^t}{\partial t} = [X^t, P],$$

compare with section “[Preliminaries from Differential Geometry](#)”. In fact, exactly this choice  $h = P_1$  was taken by Roussarie (1987), also see the proof of Takens (1974). Notice that  $\Phi = h_N \circ h_{N-1} \circ \dots \circ h_3 \circ h_2$ . Moreover notice that if the vector field  $P$  is in a given Lie-algebra of vector fields, its time 1 map  $P_1$  is in the corresponding Lie-group. In particular, if  $P$  is Hamiltonian,  $P_t$  is canonical or symplectic, and so on.

For the validity of this set-up for a more general Lie-subalgebra of the Lie-algebra of all  $C^\infty$  vector fields, one has to study how far the *grading*

$$\prod_{m=1}^{\infty} H^m(\mathbb{R}^n)$$

of the formal power series, as well as the *splittings*

$$B^m \oplus G^m = H^m(\mathbb{R}^n),$$

are compatible with the Lie-algebra at hand. This issue was addressed by Broer (1981, 1993) axiomatically, in terms of graded and filtered Lie-algebras. Moreover, the methods concerning the choice of  $G^m$  briefly mentioned at the end of the previous section, all carry over to a more general Lie-algebra set-up. As a consequence there exists a version of Theorem 6 in this setting. Instead of pursuing this further, we discuss its implications in a few relevant settings.

### The Volume Preserving and Symplectic Case

On  $\mathbb{R}^n$ , resp.  $\mathbb{R}^{2n}$ , we consider a volume form or a symplectic form, both denoted by  $\sigma$ . We assume, that

$$\sigma = dx_1 \wedge \cdots \wedge dx_n, \quad \text{resp.}$$

$$\sigma = \sum_{j=1}^n dx_j \wedge dx_{j+n}.$$

In both cases, let  $\mathcal{X}_\sigma$  denote the Lie-algebra of  $\sigma$ -preserving vector fields, i.e., vector fields  $X$  such that  $\mathcal{L}_X \sigma = 0$ . Here  $\mathcal{L}$  again denotes the Lie-derivative, see section “[The Normal Form Procedure](#)”.

Indeed, one defines

$$\begin{aligned} \mathcal{L}_X \sigma(x) &= \frac{d}{dt} \Big|_{t=0} (X_t)^* \sigma(x) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \{ (X_t)^* \sigma(x) - \sigma(x) \}. \end{aligned}$$

Properties, similar to Proposition 5, hold here. Since in both cases  $s$  is a closed form, one shows by ‘the magic formula’ (Spivak 1970) that

$$\begin{aligned} \mathcal{L}_X \sigma(x) &= d(\iota_X \sigma) + \iota_X d\sigma \\ &= d(\iota_X \sigma). \end{aligned}$$

Here  $\iota_X \sigma$  denotes the flux-operator defined by  $\iota_X \sigma(Y) = \sigma(X, Y)$ . In the volume-preserving case the latter expression denotes  $\text{div}(X)\sigma$  and we see that preservation of  $s$  exactly means that  $\text{div}(X) = 0$ : the divergence of  $X$  vanishes. In the Hamiltonian case we conclude that the 1-form  $\iota_X \sigma$  is closed and hence (locally) of the form  $dH$ , for a Hamilton function  $H$ . In both cases, the fact that for a transformation  $h$  the fact that  $h^* \sigma = \sigma$  implies that with  $X$  also  $h_* X$  is  $s$ -preserving. Moreover, for a  $s$ -preserving vector field  $P$  and  $h = P_1$  one can show that indeed  $h^* \sigma = \sigma$ .

One other observation is, that by the *homogeneity* of the above expressions for  $s$ , the homogeneous parts of the Taylor series of  $s$ -preserving vector fields are again  $s$ -preserving. This exactly means that

$$H^m(\mathbb{R}^{(2)n}) \cap \mathcal{X}_\sigma,$$

$m = 1, 2, \dots$ , grades the formal power series corresponding to  $\mathcal{X}_\sigma$ . Here, notice that

$$H^1(\mathbb{R}^{(2)n}) \cap \mathcal{X}_\sigma = \mathfrak{sl}(n, \mathbb{R}), \quad \text{resp.} \quad \mathfrak{sp}(2n, \mathbb{R}),$$

the *special*- resp. the *symplectic* linear algebra.

In summary we conclude that both the symplectic and the volume preserving setting are covered by the axiomatic approach of (Broer 1981, 1993) and that an appropriate version of Theorem 6 holds here. Below we shall illustrate this with a few examples.

### External Parameters

A  $C^\infty$  family  $X = X^\lambda(x)$  of vector fields on  $\mathbb{R}^n$ , with a multi-parameter  $\lambda \in \mathbb{R}^p$ , can be regarded as one  $C^\infty$  vector field on the product space  $\mathbb{R}^n \times \mathbb{R}^p$ . Such a vector field is *vertical*, in the sense that it has no components in the  $\lambda$ -direction. In other words, if  $\pi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^p$  is the natural projection on the second component,  $X$  is tangent to the fibers of  $\pi$ . It is easily seen that this property defines a Lie-subalgebra of the Lie-algebra of all  $C^\infty$  vector fields on  $\mathbb{R}^n \times \mathbb{R}^p$ . Again, by the linearity of this projection, the gradings and splittings are compatible. The normal form transformations  $\Phi$  preserve the parameter  $\lambda$ , i.e.,  $\Phi \circ \pi = \pi$ .

When studying a bifurcation problem, we often consider systems  $X = X^\lambda(x)$  locally defined near  $(x, \lambda) = (0, 0)$ , considering series expansions both in  $x$  and in  $\lambda$ . Then, in the  $N$ th order normalization  $\Phi_* X$ , the normalized part consists of a polynomial in  $y$  and  $\lambda$ , while the remainder term is of the form  $O(|y|^{N+1} + |\lambda|^{N+1})$ .

As in the previous case, we shall not formulate the present analogue of Theorem 6 for this case, but illustrate its meaning in examples.

### The Reversible Case

In the reversible case a linear involution  $R$  is given, while for the vector fields we require  $R_* X = -X$ . Let  $\mathcal{X}_R$  denote the class of all such reversible vector fields. Also, let  $\mathcal{C}$  denote the class of all  $X$  such that  $R_* X = X$ . Then, both  $\mathcal{X}_R$  and  $\mathcal{C}$  are linear spaces of vector fields. Moreover,  $\mathcal{C}$  is a Lie-subalgebra. Associated to  $\mathcal{C}$  is the group of diffeomorphisms that commute with  $R$ , i.e., the *R-equivariant* transformations. Also it is easy to see that for each of these diffeomorphisms  $F$  one has  $\Phi_*(\mathcal{X}_R) \subset \mathcal{X}_R$ . The above approach applies to this situation in a straightforward manner. The gradings and splittings fit, while we have to choose the infinitesimal generator  $P$  from the set  $\mathcal{C}$ . For details compare with (Broer et al. 2007d).

**Remark** In the case with parameters, it sometimes is possible to obtain an alternative normal form where the normalized part is polynomial in  $y$  alone, with coefficients that depend smoothly on  $\lambda$ . A necessary condition for this is that the origin  $y = 0$  is an equilibrium for all values of  $\lambda$  in some neighborhood of 0. To be precise, at the  $N$ th order we can achieve smooth dependence of the coefficients on  $\lambda$  for  $\lambda \in \Lambda_N$ , where  $\Lambda_N$  is a neighborhood of  $\lambda = 0$ , that may shrink to  $\{0\}$  as  $N \rightarrow \infty$ . So, for  $N \rightarrow \infty$  only the formal aspect remains, as is the case in the above approach. This alternative normal form can be obtained by a proper use of the Implicit Function Theorem in the spaces  $H^m(\mathbb{R}^n)$ ; for details e.g., see Sect. 2.2 in Vanderbauwhede (1989). For another discussion on this topic, cf. Section 7.6.2 in Dumortier et al. (Broer et al. 1991).

In section “The Normal Form Procedure” the role of symmetry was considered regarding the semisimple part of the matrix  $A$ . A question is how this discussion generalizes to Lie-subalgebra’s of vector fields.

**Example (Volume preserving, parameter dependent axial symmetry (Broer 1981, 1993))** On  $\mathbb{R}^3$  consider a 1-parameter family  $X^\lambda$  of vector fields, preserving the standard volume  $\sigma = dx_1 \wedge dx_2 \wedge dx_3$ . Assume that  $X^0(0) = 0$  while the spectrum of  $D_0 X^0$  consists of the eigenvalues  $\pm i$  and 0. For the moment regarding  $\lambda$  as an extra state space coordinate, we obtain a vertical vector field on  $\mathbb{R}^4$  and we apply a combination of the above considerations. The ‘generic’ Jordan normal form then is

$$A = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (2)$$

with an obvious splitting in semisimple and nilpotent part. The considerations of section “The Normal Form Procedure” then directly apply to this situation. For any  $N$  this yields a transformation  $\Phi : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ , with  $\Phi(0) = 0$ , preserving both the projection to the 1-dimensional parameter space

and the volume of the 3-dimensional phase space, such that the normalized,  $N$ th order part of  $\Phi_* X(y, \lambda)$ , in cylindrical coordinates  $y_1 = r \cos \varphi$ ,  $y_2 = r \sin \varphi$ ,  $y_3 = z$ , has the rotationally symmetric form

$$\begin{aligned} \dot{\varphi} &= f(r^2, z, \lambda) \\ \dot{r} &= rg(r^2, z, \lambda) \\ \dot{z} &= h(r^2, z, \lambda), \end{aligned}$$

again, for suitable polynomials  $f$ ,  $g$  and  $h$ . Note, that in cylindrical coordinates the volume has the form  $\sigma = r dr \wedge d\varphi \wedge dz$ . Again the functions  $f$ ,  $g$  and  $h$  have to fit with the linear part. In particular we find that  $h(r^2, z, \lambda) = \lambda + az + \dots$ , observing that for  $\lambda \neq 0$  the origin is no equilibrium point.

**Remark** If  $A = A_s + A_n$  is the canonical splitting of  $A$  in  $H^1(\mathbb{R}^n) = \mathfrak{gl}(n, \mathbb{R})$ , then automatically both  $A_s$  and  $A_n$  are in the subalgebra under consideration. In the volume preserving setting this can be seen directly. In general the same holds true as soon as the corresponding linear Lie-group is algebraic, see (Broer 1981, 1993) and the references given there.

### The Hamiltonian Case

The Normal Form Theory in the Hamiltonian case goes back at least to Poincaré (1928) and Birkhoff (1927). Other references are, for instance, Gustavson (Gustavson 1966), Arnold (1980, 1983, 1988), Sanders, Verhulst and Murdock (1985), Van der Meer (1985), Broer, Chow and Kim (1993). In section “Preservation of Structure” we already saw that the axiomatic Lie algebra approach of (Broer 1981, 1993) applies here, especially since the symplectic group  $\mathrm{SP}(2n, \mathbb{R})$  is algebraic. The canonical form here usually goes with the name Williamson, compare Galin (1982), Koçak (1984) and Hoveijn (1996). We discuss how the Lie algebra approach compares to the literature.

The Lie algebra of Hamiltonian vector fields can be associated to the Poisson-algebra of Hamilton functions as follows, even in an arbitrary symplectic setting. As before, the symplectic form is denoted by  $s$ . We recall, that for any Hamiltonian  $H$  the corresponding Hamiltonian

vector field  $X_H$  is given by  $dH = \sigma(X_H, \cdot)$ . Now, let  $H$  and  $K$  be Hamilton functions with corresponding vector fields  $X_H$  resp.  $X_K$ . Then

$$X_{\{H, K\}} = [X_H, X_K],$$

implying that the map  $H \mapsto X_H$  is a morphism of Lie algebra's. By definition this map is surjective, while its kernel consists of the (locally) constant functions.

This implies, that the normal form procedure can be completely rephrased in terms of the Poisson-bracket. We shall now demonstrate this by an example, similar to the previous one.

**Example (Symplectic parameter dependent rotational symmetry (Broer et al. 1993))** Consider  $\mathbb{R}^4$  with coordinates  $(x_1, y_1, x_2, y_2)$  and the standard symplectic form  $\sigma = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$ , considering a  $C^\infty$  family of Hamiltonian functions  $H^\lambda$ , where  $\lambda \in \mathbb{R}$  is a parameter. The fact that  $dH = \sigma(X_H, \cdot)$  in coordinates means  $\dot{x}_j = \frac{\partial}{\partial y_j} H$ ,  $\dot{y}_j = -\frac{\partial}{\partial x_j} H$ , for  $j = 1, 2$ . We assume that for  $\lambda = 0$  the origin of  $\mathbb{R}^4$  is a singularity. Then we expand as a Taylor series in  $(x, y, \lambda)$

$$H^\lambda(x, y) = H_2(x, y, \lambda) + H_3(x, y, \lambda) + \dots,$$

where the  $H_m$  is homogeneous of degree  $m$  in  $(x, y, \lambda)$ . It follows for the corresponding Hamiltonian vector fields that

$$X_{H_m} \in H^{m-1}(\mathbb{R}^4),$$

in particular  $X_{H_2} \in \mathfrak{sp}(4, \mathbb{R}) \subset H^1(\mathbb{R}^4)$ . Let us assume that this linear part  $X_{H_2}$  at  $(x, y, \lambda) = (0, 0, 0)$  has eigenvalues  $\pm i$  and a double eigenvalue 0. One 'generic' Williamson's normal form then is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3)$$

compare with (2), corresponding to the quadratic Hamilton function

$$H_2(x, y, \lambda) = I + \frac{1}{2}y_2^2,$$

where  $I := \frac{1}{2}(x_1^2 + y_1^2)$ . It is straightforward to give the generic matrix for the linear part in the extended state space  $\mathbb{R}^5$ , see above, so we will leave this to the reader.

The semisimple part  $A_s = X_I$  now can be used to obtain a rotationally symmetric normal form, as before. In fact, for any  $N \in \mathbb{N}$  there exists a canonical transformation  $\Phi(x, y, \lambda)$ , which keeps the parameter fixed, and a polynomial  $F(I, x_2, y_2, \lambda)$ , such that

$$(H^\circ \Phi^{-1})(x, y, \lambda) = F(I, x_2, y_2, \lambda) + O(I + x_2^2 + y_2^2 + \lambda^2)^{(N+1)/2}.$$

Instead of using the adjoint action  $\text{ad}_A$  on the spaces  $H^{m-1}(\mathbb{R}^5)$ , we also may use the adjoint action

$$\text{ad}_{H_2} : f \mapsto \{H_2, f\},$$

where  $f$  is a polynomial function of degree  $m$  in  $(x_1, y_1, x_2, y_2, \lambda)$ . In the vector field language, we choose the 'good' space  $G^{m-1} \subset \ker \text{ad}_{m-1} A_s$ , which, in the function language, translates to a 'good' subset of  $\ker \text{ad}_I$ .

Whatsoever, the normalized part of the Hamilton function  $H^\circ \Phi^{-1}$ , viz. the vector field  $\Phi_* X_H = X_{H^\circ \Phi^{-1}}$ , is rotationally symmetric. The fact that the Hamilton function  $F$  Poisson-commutes with  $I$  exactly amounts to invariance under the action generated by the vector field  $X_I$ , in turn implying that  $I$  is an integral of  $X_F$ . Indeed, if we define a  $2\pi$ -periodic variable  $\varphi$  as follows:

$$x_1 = \sqrt{2I} \sin \varphi, \quad y_1 = \sqrt{2I} \cos \varphi,$$

then  $\sigma = dI \wedge d\varphi + dx_2 \wedge dy_2$ , implying that the normalized vector field  $X_F$  has the canonical form

$$\begin{aligned} \dot{I} &= 0, \quad \dot{\varphi} = -\frac{\partial F}{\partial I}, \\ \dot{x}_2 &= \frac{\partial F}{\partial y_2}, \quad \dot{y}_2 = -\frac{\partial F}{\partial x_2}. \end{aligned}$$

Notice that  $\varphi$  is a cyclic variable, making the fact that  $I$  is an integral clearly visible. Also

observe that  $\dot{\varphi} = -1 + \dots$ . As before, and as in, e.g., the central force problem, this enables a reduction to lower dimension. Here, the latter two equations constitute the *reduction to 1 degree of freedom*: it is a family of planar Hamiltonian vector fields, parametrized by  $I$  and  $\lambda$ .

**Remark**

- This example has many variations. First of all it can be simplified by omitting parameters and even the zero eigenvalues. The conclusion then is that a planar Hamilton function with a non-degenerate minimum or maximum has a formal rotational symmetry, up to canonical coordinate changes.
- It also can be easily made more complicated, compare with Proposition 7 in section “[The Normal Form Procedure](#)”. Given an equilibrium with purely imaginary eigenvalues  $\pm i \omega_1, \pm i \omega_2, \dots, \pm i \omega_m$  and with  $2(n - m)$  zero eigenvalues. Provided that there are no resonances up to order  $N + 1$ , see (1), also here we conclude that Hamiltonian truncation at the order  $N$ , by canonical transformations can be given the form

$$F(I_1, \dots, I_m, x_{m+1}, y_{m+1}, \dots, x_n, y_n),$$

where  $I_j := \frac{1}{2}(x_j^2 + y_j^2)$ . As in Proposition 7 this normal form has a toroidal symmetry. Writing  $x_j = \sqrt{2I_j} \sin \varphi_j, y_j = \sqrt{2I_j} \cos \varphi_j$ , we obtain the canonical system of equations

$$\begin{aligned} \dot{I}_j &= 0, & \dot{\varphi}_j &= -\frac{\partial F}{\partial I_j}, \\ \dot{x}_\ell &= \frac{\partial F}{\partial y_\ell}, & \dot{y}_\ell &= -\frac{\partial F}{\partial x_\ell}, \end{aligned}$$

$1 \leq j \leq m, \quad m + 1 \leq l \leq n$ . Note that  $\dot{\varphi}_j = -\omega_j + \dots$ . In the case  $m = n$  we deal with an elliptic equilibrium. The corresponding result usually is named the Birkhoff normal form (Arnold 1980; Birkhoff 1927; Gustavson 1966), ▶ “[Dynamics of Hamiltonian Systems](#)”. Then, the variables  $(I_j, \varphi_j)$ ,  $1 \leq j \leq m$ , are a set of *action-angle variables* for the truncated part of order  $N$ , (Arnold 1980).

- Returning to algorithmic issues, in addition to subsection “[On the Choices of the Complementary Space and of the Normalizing Transformation](#)”, we give a few further references in cases where a structure is being preserved: Broer (1981), Deprit (1969, 1981), Meyer (1984) and Ferrer, Hanßmann, Palacián and Yanguas (2002). It is to be noted that this kind of Hamiltonian result also can be obtained, where the coordinate changes are constructed using *generating functions*, compare with, e.g., (Arnold 1980, 1988; Siegel and Moser 1971), see also the discussion in Broer, Hoveijn, Lunter and Vegter (1998a) and in particular Sect. 4.4.2 in Broer et al. (2003a). For another, computationally very effective normal form algorithm, see Giorgilli and Galgani (1978). Also compare with references therein.

## Semi-local Normalization

This section roughly consists of two parts. To begin with, a number of subsections are devoted to related formal normal form results, nearfixed points of diffeomorphisms and near periodic solutions and invariant tori of vector fields.

### A Diffeomorphism Near a Fixed Point

We start by formulating a result by Takens:

**Theorem 8 (Takens normal form (Takens 1973))** Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a  $C^\infty$  diffeomorphism with  $T(0) = 0$  and with a canonical decomposition of the derivative  $D_0T = S + N$  in semisimple resp. nilpotent part. Also, let  $N \in \mathbb{N}$  be given, then there exists diffeomorphism  $\Phi$  and a vector field  $X$ , both of class  $C^\infty$ , such that  $S_*X = X$  and

$$\Phi^{-1} \circ T \circ \Phi = S^*X_1 + O(|y|^{N+1}).$$

Here, as before,  $X_1$  denotes the flow over time 1 of the vector field  $X$ . Observe that the vector field  $X$  necessarily has the origin as an equilibrium point. Moreover, since  $S_*X = X$ , the vector field  $X$  is invariant with respect to the group generated by  $S$ .

The proof is a bit more involved than Theorem 6 in section “[The Normal Form Procedure](#)”, but it has the same spirit, also compare with (Chow et al. 1994). In fact, the Taylor series of  $T$  is modified step-by-step, using coordinate changes generated by homogeneous vector fields of the proper degree.

After a reduction to center manifolds the spectrum of  $S$  is on the complex unit circle and Theorem 8 especially is of interest in cases where this spectrum consists of roots of unity, i.e., in the case of resonance. Compare with (Broer and Vegter 2008; Chow et al. 1994; Takens 1973).

Again, the result is completely phrased in terms of Lie algebra's and groups and therefore bears generalization to many contexts with a preserved structure, compare section “[Preservation of Structure](#)”. The normalizing transformations of the induction process then are generated from the corresponding Lie algebra. For a symplectic analogue see Moser (1968). Also, both in Broer, Chow and Kim (1993) and in Broer and Vegter (1992), symplectic cases with parameters are discussed, where  $S = \text{Id}$ , the Identity Map, resp.  $S = -\text{Id}$ , the latter involving a period-doubling bifurcation.

**Remark** Let us consider a symplectic map  $T$  of the plane, which means that  $T$  preserves both area and orientation. Assume that  $T$  is fixing the origin, while the eigenvalues of  $S = D_0T$  are on the unit circle, without being roots of unity. Then  $S$  generates the rotation group  $\text{SO}(2, \mathbb{R})$ , so the vector field  $X$ , which has divergence zero, in this case is rotationally symmetric. Again this result often goes with the name of Birkhoff.

### Near a Periodic Solution

The Normal Form Theory at a periodic solution or closed orbit has a lot of resemblance to the local theory we met before.

To fix thoughts, let us consider a  $C^\infty$  vector field of the form

$$\begin{aligned}\dot{x} &= f(x, y) \\ \dot{y} &= g(x, y),\end{aligned}\tag{4}$$

with  $(x, y) \in \mathbb{T}^1 \times \mathbb{R}^n$ . Here  $\mathbb{T}^1 = \mathbb{R}/(2\pi\mathbb{Z})$ . Assuming  $y = 0$  to be a closed orbit, we consider the formal Taylor series with respect to  $y$ , with

$x$ -periodic coefficients. By Floquet Theory, we can assume that the coordinates  $(x, y)$  are such that

$$f(x, y) = \omega + O(|y|), \quad g(x, y) = \Omega y + O(|y|^2),$$

where  $\omega \in \mathbb{R}$  is the frequency of the closed orbit and  $\Omega \in \mathfrak{gl}(n, \mathbb{R})$  its Floquet matrix. Again, the idea is to ‘simplify’ this series further. To this purpose we introduce a grading as before, letting  $H^m = H^m(\mathbb{T}^1 \times \mathbb{R}^n)$  be the space of vector fields

$$Y(x, y) = L(x, y) \frac{\partial}{\partial x} + \sum_{j=1}^n M_j(x, y) \frac{\partial}{\partial y_j},$$

with  $L(x, y)$  and  $M_j(x, y)$  homogeneous in  $y$  of degree  $m - 1$  resp.  $m$ . Notice, that this space  $H^m$  is infinite-dimensional. However, this is not at all problematic for the things we are doing here. By this definition, we have that

$$A := \omega \frac{\partial}{\partial x} + \Omega y \frac{\partial}{\partial y}$$

is a member of  $H^1$  and with this normally linear part we can define an adjoint representation  $\text{ad}A$  as before, together with linear maps

$$\text{ad}_m A : H^m \rightarrow H^m.$$

Again we assume to have a decomposition

$$G^m \oplus \text{Im}(\text{ad}_m A) = H^m,$$

where the aim is to transform the terms of the series successively into the  $G^m$ , for  $m = 2, 3, 4, \dots$

The story now runs as before. In fact, the proof of Theorem 6 in section “[The Normal Form Procedure](#)”, as well as its Lie algebra versions indicated in section “[Preservation of Structure](#)”, can be repeated almost verbatim for this case. Moreover, if  $\Omega = \Omega_s + \Omega_n$  is the canonical splitting in semisimple and nilpotent part, then

$$\omega \frac{\partial}{\partial x} + \Omega_s y \frac{\partial}{\partial y}$$

gives the semisimple part of  $\text{ad}_m A$ , as can be checked by a direct computation. From this



computation one also deduces the non-resonance conditions needed for the present torus-symmetric analogue of Proposition 7 in section “[The Normal Form Procedure](#)”.

There are different cases *with* resonance either between the imaginary parts of the eigenvalues of  $\Omega$  (normal resonance) or between the latter and the frequency  $\omega$  (normal-internal resonance). All of this extends to the various settings with preservation of structure as discussed before. In all cases direct analogues of the Theorems 6 and 8 are valid. General references in this direction are Arnold (1983, 1988), Bruno (1989, 2000), Chow, Li and Wang (1994), Iooss (1988), Murdock (2003) or Sanders, Verhulst and Murdock (1985).

### Remark

- This approach also is important for non-autonomous systems with periodic time dependence. Here the normalization procedure includes averaging. As a special case of the above form, we obtain a system

$$\begin{aligned}\dot{x} &= \omega \\ \dot{y} &= g(x, y),\end{aligned}$$

so where  $x \in \mathbb{T}^1$  is proportional to the time. Apart from the general references given above, we also refer to, e. g., Broer and Vegter (1992) and to Broer, Roussarie and Simó (2001; Broer et al. 1996b). The latter two applications also contain parameters and deal with bifurcations. Also compare with Verhulst ▶ “[Perturbation Analysis of Parametric Resonance](#)”.

- A geometric tool that can be successfully applied in various resonant cases is a *covering space*, obtained by a Van der Pol transformation (or by passing to co-rotating coordinates). This involves equivariance with respect to the corresponding deck group. This setting (with or without preservation of structure) is completely covered by the general Lie algebra approach as described above.

For the Poincaré map this deck group symmetry directly yields the normal form symmetry of

Theorem 8. For applications in various settings, with or without preservation of structure, see (Broer and Vegter 1992, 2008; Broer et al. 2003b). This normalization technique is effective for studying bifurcation of subharmonic solutions. In the case of period doubling the covering space is just a double cover. In many cases Singularity Theory turns out to be useful.

### Near a Quasi-periodic Torus

The approach of the previous subsection also applies at an invariant torus, provided that certain requirements are met. Here we refer to Braaksma, Broer and Huitema (1990), Broer and Takens (1989) and Bruno (1989; Bruno 2000).

Let us consider a  $C^\infty$ -system

$$\begin{aligned}\dot{x} &= f(x, y) \\ \dot{y} &= g(x, y)\end{aligned}\tag{5}$$

as before, with  $(x, y) \in \mathbb{T}^m \times \mathbb{R}^n$ . Here  $\mathbb{T}^m = \mathbb{R}^m / (2\pi\mathbb{Z})^m$ . We assume that  $f(x, y) = \omega + O(|y|)$ , which implies that  $y = 0$  is an invariant  $m$ -torus, with on it a constant vector field with frequency-vector  $\omega$ . We also assume that  $g(x, y) = \Omega y + O(|y|^2)$ , which is the present analogue of the Floquet form as known in the periodic case, with  $\Omega \in \mathfrak{gl}(n, \mathbb{R})$ , independent of  $x$ . Contrarius to the situations for  $m = 1$  and  $n = 1$ , in general reducibility to Floquet form is not possible. Compare with (Broer and Sevryuk 2008; Broer et al. 1990, 1996a; Ciocci et al. 2005) and references therein. For a similar approach of a system that is not reducible to Floquet form, compare with (Broer et al. 1999).

Presently we assume this reducibility, expanding in formal series with respect to the variables  $y$ , where the coefficients are functions on  $\mathbb{T}^m$ . These coefficients, in turn, can be expanded in Fourier series. The aim then is, to ‘simplify’ this combined series by successive coordinate changes, following the above procedure. As a second requirement it then is needed that certain *Diophantine* conditions are satisfied on the pair  $(\omega, \Omega)$ . Below we give more details on this. Instead of giving general results we again refer to (Broer and Takens 1989; Broer et al. 1990, 2007a, d; Bruno 1989, 2000). Moreover, to fix thoughts, we



present a simple example with a parameter. Here again a direct link with averaging holds.

**Example (Toroidal symmetry with small divisors (Ciocci et al. 2005))** Given is a family of vector fields  $\dot{x} = X(x, \lambda)$  with  $(x, \lambda) \in \mathbb{T}^m \times \mathbb{R}$ . We assume that  $X = X(x, \lambda)$  has the form

$$X(x, \lambda) = \omega + f(x, \lambda),$$

with  $f(x, 0) \equiv 0$ . It is assumed that the frequency vector  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  has components that satisfy the Diophantine non-resonance condition

$$|\langle \omega, k \rangle| \geq \gamma |k|^{-\tau}, \quad (6)$$

for all  $k \in \mathbb{Z} \setminus \{0\}$ . Here  $\gamma > 0$  and  $\tau > n - 1$  are prescribed constants. We note that  $\tau > n - 1$  implies that this condition in  $\mathbb{R}^n = \{\omega\}$  excludes a set of Lebesgue measure  $O(\gamma)$  as  $\gamma \downarrow 0$ , compare with (Ciocci et al. 2005). It follows, that by successive transformations of the form

$$h : (x, \lambda) \mapsto (x + P(x, \lambda), \lambda)$$

the  $x$ -dependence of  $X$  can be pushed away to ever higher order in  $\lambda$ , leading to a formal normal form

$$\dot{\xi} = \omega + g(\lambda),$$

with  $g(0) = 0$ . Observe that in this case ‘simple’ means  $x$ - (or  $\xi$ -) independent. Therefore, in a proper formalism,  $x$ -independent systems constitute the spaces  $G^m$ . Indeed, in the induction process we get

$$X(x, \lambda) = \omega + g_2(\lambda) + \dots + g_{N-1}(\lambda) + f_N(x, \lambda) + O(|\lambda|^{N+1}),$$

compare the proof of Theorem 6 in section “[The Normal Form Procedure](#)”. Writing  $\xi = x + P(x, \lambda)$ , with  $P(x, \lambda) = \tilde{P}(x, \lambda)\lambda^N$ , we substitute

$$\begin{aligned} \dot{\xi} &= (\text{Id} + D_\xi P)\dot{x} \\ &= \omega + g_2(\lambda) + \dots + g_{N-1}(\lambda) \\ &\quad + f_N(x, \lambda) + D_x P(x, \lambda)\omega + O(|\lambda|^{N+1}), \end{aligned}$$

Where we express the right-hand side in  $x$ . So we have to satisfy an equation

$$D_x P(x, \lambda)\omega + f_N(x, \lambda) \equiv c\lambda^N \bmod O(|\lambda|^{N+1}),$$

for a suitable constant  $c$ . Writing  $f_N(x, \lambda) = \tilde{f}_N(x, \lambda)\lambda^N$ , this amounts to

$$D_x \tilde{P}(x, \lambda)\omega = -\tilde{f}_N(x, \lambda) + c,$$

which is the present form of the homological equation. If

$$\tilde{f}_N(x, \lambda) = \sum_{k \in \mathbb{Z}^n} a_k(\lambda) e^{i\langle x, k \rangle},$$

then  $c = a_0$ , i.e., the  $m$ -torus average

$$a_0(\lambda) = \frac{1}{(2\pi)^m} \int_{\mathbb{T}^m} \tilde{f}_N(x, \lambda) dx.$$

Moreover,

$$\tilde{P}(x, \lambda) = \sum_{k \neq 0} \frac{a_k(\lambda)}{i\langle \omega, k \rangle} e^{i\langle x, k \rangle}.$$

This procedure formally only makes sense if the frequencies  $(\omega_1, \omega_2, \dots, \omega_m)$  have no resonances, which means that for  $k \neq 0$  also  $\langle \omega, k \rangle \neq 0$ . In other words this means that the components of the frequency vector  $\omega$  are rationally independent. Even then, the denominator  $i\langle \omega, k \rangle$  can become arbitrarily small, so casting doubt on the convergence. This problem of *small divisors* is resolved by the Diophantine conditions (6). For further reference, e.g., see (Arnold 1983; Broer and Sevryuk 2008; Broer et al. 1990, 1996a; Ciocci et al. 2005): for real analytic  $X$ , by the Paley–Wiener estimate on the exponential decay of the Fourier coefficients, the solution  $P$  again is real analytic. Also in the  $C^\infty$ -case the situation is rather simple, since then the coefficients in both cases decay faster than any polynomial.

#### Remark

- The discussion at the end of subsection “[Near a Periodic Solution](#)” concerning normalization

at a periodic solution, largely extends to the present case of a quasi-periodic torus. Assuming reducibility to Floquet form, we now have a normally linear part

$$\omega \frac{\partial}{\partial x} + \Omega y \frac{\partial}{\partial y},$$

with a frequency vector  $\omega \in \mathbb{R}^n$ . As before  $\Omega \in \mathfrak{gl}(m, \mathbb{R})$ . For the corresponding KAM Perturbation Theory the Diophantine conditions (6) on the frequencies are extended by including the normal frequencies, i.e., the imaginary parts  $\omega_1^N, \dots, \omega_s^N$  of the eigenvalues of  $\Omega$ . To be precise, for  $\gamma > 0$  and  $\tau > n - 1$ , above the conditions (6), these extra Melnikov conditions are given by

$$\begin{aligned} |\langle \omega, k \rangle + \omega_j^N| &\geq \gamma |k|^{-\tau} \\ |\langle \omega, k \rangle + 2\omega_j^N| &\geq \gamma |k|^{-\tau} \\ |\langle \omega, k \rangle + \omega_j^N + \omega_\ell^N| &\geq \gamma |k|^{-\tau}, \end{aligned} \quad (7)$$

for all  $k \in \mathbb{Z}^n \setminus \{0\}$  and for  $j, \ell = 1, 2, \dots, s$  with  $\ell \neq j$ . See below for the description of an application to KAM Theory (and more references) in the Hamiltonian case.

- In this setting normal resonances can occur between the  $\omega_j^N$  and normal-internal resonances between the  $\omega_j^N$  and  $\omega$ . Certain strong normal-internal resonances occur when one of the left-hand sides of (7) vanishes. For an example see below. Apart from these now also internal resonances between the components of  $\omega$  come into play. The latter generally lead to destruction of the invariant torus.
- As in the periodic case also here covering spaces turn out to be useful for studying various resonant bifurcation scenarios often involving applications of both Singularity Theory and KAM Theory. Compare with (Broer et al. 1990, 2003c, 2008, n.d.).

### Non-formal Aspects

Up to this moment (almost) all considerations have been formal, i.e., in terms of formal power series. In general, the Taylor series of a  $C^\infty$ -function, say,  $\mathbb{R}^n \rightarrow \mathbb{R}$  will be divergent. On the other

hand, any formal power series in  $n$  variables occurs as the Taylor series of some  $C^\infty$ -function. This is the content of a theorem by É. Borel, cf. Narasimhan (Narasimhan 1968).

We briefly discuss a few aspects regarding convergence or divergence of the normalizing transformation or of the normalized series. We recall that the growth rate of the formal series, including the convergent case, is described by the Gevrey index, compare with, e.g., (Baldomá and Haro 2008; Marco and Sauzin 2002; Ramis 1994; Ramis and Schäfke 1996).

### Normal Form Symmetry and Genericity

For the moment we assume that all systems are of class  $C^\infty$ . As we have seen, if the normalization procedure is carried out to some finite order  $N$ , the transformation  $\Phi$  is a real analytic map. If we take the limit for  $N \rightarrow \infty$ , we only get formal power series  $\Phi$ , but, by the Borel Theorem, a ‘real’  $C^\infty$ -map  $F$  exists with  $\Phi$  as its Taylor series.

Let us discuss the consequences of this, say, in the case of Proposition 7 in section “The Normal Form Procedure”. Assuming that there are no resonances at all between the  $\omega_j$ , as a corollary, we find a  $C^\infty$ -map  $y = \Phi(x)$  and a  $C^\infty$  vector field  $p = p(y)$ , such that:

- The vector field  $\Phi_* X - p$ , in corresponding generalized cylindrical coordinates has the symmetric  $C^\infty$ -form

$$\begin{aligned} \dot{\varphi}_j &= f_j(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n) \\ \dot{r}_j &= r_j g_j(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n) \\ \dot{z}_\ell &= h_\ell(r_1^2, \dots, r_m^2, z_{n-2m+1}, \dots, z_n), \end{aligned}$$

where  $f_j(0) = \omega_j$  and  $h_\ell(0) = 0$  for  $1 \leq j \leq m$ ,  $n - 2m + 1 \leq \ell \leq n$ .

- The Taylor series of  $p$  identically vanishes at  $y = 0$ .

Note, that an  $\infty$ -ly flat term  $p$  can have component functions like  $e^{-1/y_1^2}$ . We see, that the  $m$ -torus symmetry only holds up to such flat terms. Therefore, this symmetry, if present at all, also can be destroyed again by a *generic* flat

‘perturbation’. We refer to Broer and Takens (1989), and references therein, for further consequences of this idea. The main point is, that by a Kupka–Smale argument, which generically forbids so much symmetry, compare with (Palis and de Melo 1982).

**Remark** The Borel Theorem also can be used in the reversible, the Hamiltonian and the volume preserving setting. In the latter two cases we exploit the fact that a structure preserving vector field is generated by a function, resp. an  $(n - 2)$ -form. Similarly the structure preserving transformations have such a generator. On these generators we then apply the Borel Theorem. Many Lie algebra’s of vector fields have this ‘Borel Property’, saying that a formal power series of a transformation can be represented by a  $C^\infty$  map in the same structure preserving setting.

### On Convergence

The above topological ideas also can be pursued in many real analytic cases, where they imply a generic *divergence* of the normalizing transformation. For an example in the case of the HamiltonianBirkhof normal form (Birkhoff 1927; Gustavson 1966; Siegel and Moser 1971) compare with Broer and Tangerman (Broer and Tangerman 1986) and its references. As an example we now deal with the linearization of a holomorphic germ in the spirit of section “The Normal Form Procedure”.

**Example (Holomorphic linearization (Bruno 1965; Milnor 2006; Yoccoz 1995))** A holomorphic case concerns the linearization of a local holomorphic map  $F : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$  of the form  $F(z) = \lambda z + f(z)$  with  $f(0) = f'(0) = 0$ . The question is whether there exists a local biholomorphic transformation  $\Phi : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$  such that

$$\Phi \circ F = \lambda \cdot \Phi.$$

We say that  $\Phi$  linearizes  $F$  near its fixed point 0. A formal solution as in section “The Normal Form Procedure” generally exists for all  $\lambda \in \mathbb{C} \setminus \{0\}$ , not equal to a root of unity. The elliptic case concerns  $\lambda$  on the complex unit circle, so of the

form  $\lambda = e^{2\pi i \alpha}$ , where  $\alpha \notin \mathbb{Q}$ , and where the approximability of  $\alpha$  by rationals is of importance, cf. Siegel (1942) and Bruno (1965). Siegel introduced a sufficient Diophantine condition related to (6), which by Bruno was replaced by a sufficient condition on the continued fraction approximation of  $\alpha$ . Later Yoccoz (1995) proved that the latter condition is both necessary and sufficient. For a description and further comments also compare with (Arnold 1983; Broer and Sevryuk 2008; Eliasson et al. 2002; Milnor 2006).

### Remark

- In certain real analytic cases Neishtadt (Neishtadt 1984), by truncating the (divergent) normal form series appropriately, obtains a remainder that is exponentially small in the perturbation parameter. Also compare with, e.g., (Arnold 1983, 1988; Baldomá and Haro 2008; Simó 1994). For an application in the context of the Bogdanov–Takens bifurcations for diffeomorphisms, see (Broer and Roussarie 2001; Broer et al. 1996b). It follows that chaotic dynamics is confined to exponentially narrow horns in the parameter space.
- The growth of the Taylor coefficients of the usually divergent series of the normal form and of the normalizing transformation can be described using Gevrey asymptotics (Baldomá and Haro 2008; Marco and Sauzin 2002; Popov 2000, 2004; Ramis 1994; Ramis and Schäfke 1996). Apart from its theoretical interest, this kind of approach is extremely useful for computational issues also compare with (Baldomá and Haro 2008; Giorgilli et al. 1989; Simó 1994, 1998, 2000).

## Applications

We present two main areas of application of the Normal Form Theory in Perturbation Theory. The former of these deals with more globally qualitative aspects of the dynamics given by normal form approximations. The latter class of applications concerns the Averaging Theorem, where the issue is that solutions remain close to approximating solutions given by a normal form truncation.

### ‘Cantorized’ Singularity Theory

We return to the discussion of the motivation in section “[Motivation](#)”, where there is a toroidal normal form symmetry up to a finite or infinite order. To begin with let us consider the quasi-periodic Hopf bifurcation (Braaksma and Broer 1987; Broer et al. 1990), which is the analogue of the Hopf bifurcation for equilibria and the Hopf–Neimark–Sacker bifurcation for periodic solutions, in the case of quasi-periodic tori. For a description comparing the differences between all these cases we refer to (Ciocci et al. 2005). For a description of the resonant dynamics in the resonant gaps see (Broer et al. 2007e) and references therein. Apart from this, a lot of related work has been done in the Hamiltonian and reversible context as well, compare with (Broer et al. 2003c, 2007a, d; Hanßmann 2007).

To be more definite, we consider families of systems defined on  $\mathbb{T}^m \times \mathbb{R}^m \times \mathbb{R}^{2p} = \{x, y, z\}$ , endowed with the symplectic form  $\sigma = \sum_{j=1}^m dx_j \wedge dy_j + dz^2$ . We start with ‘integrable’, i.e.,  $x$ -independent, families of systems of the form

$$\begin{aligned}\dot{x} &= f(y, z; \lambda) \\ \dot{y} &= g(y, z; \lambda) \\ \dot{z} &= h(y, z; \lambda),\end{aligned}\tag{8}$$

compare with (5), to be considered near an invariant  $m$ -torus  $\mathbb{T}^m \times \{y_0\} \times \{z_0\}$ , meaning that we assume  $g(y_0, z_0) = 0 = h(y_0, z_0)$ . The general interest is with the persistence of such a torus under nearly integrable perturbations of (8), where we include  $\lambda$  as a multiparameter. This problem belongs to the Parametrized KAM Theory (Broer and Sevryuk 2008; Broer et al. 1990, 1996a) of which we sketch some background now. For  $y$  near  $y_0$  and  $\lambda$  near 0 consider  $\Omega(\lambda, y) = D_z h(y, z_0; \lambda)$ , noting that  $\Omega(\lambda, y) \in \text{sp}(2p, \mathbb{R})$ . Also consider the corresponding normal linear part

$$\omega(\lambda, y) \frac{\partial}{\partial x} + \Omega(\lambda, y) z \frac{\partial}{\partial z}.$$

As a first case assume that the matrix  $\Omega(0, y_0)$  has only simple non-zero eigenvalues. Then a full

neighborhood of  $\Omega(0, y_0)$  in  $\text{sp}(2p, \mathbb{R})$  is completely parametrized by the eigenvalues of the matrices, where – in this symplectic case – we have to refrain from ‘counting double’. We roughly quote a KAM Theorem as this is known in the present circumstances (Broer et al. 1990). As a nondegeneracy condition assume that the product map

$$(\lambda, y) \mapsto (\omega(\lambda, y), \Omega(\lambda, y))$$

is a submersion near  $(\lambda, y) = (0, y_0)$ . Also assume all Diophantine conditions (6), (7) to hold. Then the parametrized system (8) is *quasi-periodically stable*, which implies persistence of the corresponding Diophantine tori near  $\mathbb{T}^m \times \{y_0\} \times \{z_0\}$  under nearly integrable perturbations.

### Remark

- A key concept in the KAM Theory is that of *Whitney smoothness* of foliations of tori over nowhere dense ‘Cantor’ sets. In real analytic cases even Gevrey regularity holds, and similarly when the original setting is Gevrey; compare with (Popov 2004; Wagener 2003). For general reference see (Broer and Sevryuk 2008; Broer et al. 1990, 1996a) and references therein.
- The gaps in the ‘Cantor sets’ are centered around the various resonances. Their union forms an open and dense set of small measure, where perturbation series diverge due to small divisors. In each gap the considerations mentioned at the end of subsections “[Near a Periodic Solution](#)” and “[Near a Quasi-Periodic Torus](#)” apply. In (Broer et al. 2003c, 2008, n. d.; Ciocci et al. 2005) the differences between period and the quasi-periodic cases are highlighted.

**Example (Quasi-periodic Hamiltonian Hopf bifurcation (Broer et al. 2007a))** As a second case we take  $p = 2$ , considering the case of normal  $1 : -1$  resonance where the eigenvalues of  $\Omega(0, y_0)$  are of the form  $\pm i \mu_0$ , for a positive  $\mu_0$ . For simplicity we only consider the non-semisimple Williamson normal form

$$\Omega(0, y_0) \sim \begin{pmatrix} 0 & -\mu_0 & 1 & 0 \\ \mu_0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\mu_0 \\ 0 & 0 & \mu_0 & 0 \end{pmatrix},$$

where  $\sim$  denotes symplectic similarity. The present format of the nondegeneracy condition regarding the product map  $(\lambda, y) \mapsto (\omega(\lambda, y), \Omega(\lambda, y))$  is as before, but now it is required that the second component  $(\lambda, y) \mapsto \Omega(\lambda, y)$  is a *versal unfolding* of the matrix  $\Omega(0, y_0)$  in  $\text{sp}(4, \mathbb{R})$  with respect to the adjoint  $\text{SP}(4, \mathbb{R})$ -action, compare with (Arnold 1983; Broer et al. 2007a, b, d; Galin 1982; Hoveijn 1996; Koçak 1984). It turns out that a standard normalization along the lines of section “[Preservation of Structure](#)” can be carried out in the  $z$ -direction, generically leading to a Hamiltonian Hopf bifurcation (van der Meer 1985), characterized by its swallowtail geometry (Thom 1989), which ‘governs’ families of invariant  $m$ -,  $(m+1)$ - and  $(m+2)$ -tori (the latter being Lagrangean). Here for the complementary spaces, see section “[The Normal Form Procedure](#)”, the  $\text{sl}(2, \mathbb{R})$ -Theory is being used (Cushman and Sanders 1986; Murdock 2003).

As before the question is what happens to this scenario when perturbing the system in a non-integrable way? In that case we need quasi-periodic Normal Form Theory, in the spirit of section “[Semi-Local Normalization](#)”. Observe that by the  $1 : -1$  resonance, difficulties occur with the third of the three Melnikov conditions (7). In a good set of Floquet coordinates the resonance can be written in the form  $\omega_j^N + \omega_\ell^N = 0$ . Nevertheless, another application of Parametrized KAM Theory (Broer and Sevryuk 2008; Broer et al. 1990, 1996a, 2007b) yields that the swallowtail geometry is largely preserved, when leaving out a dense union of resonance gaps of small measure. Here perturbation series diverge due to small divisors. What remains is a ‘Cantorized’ version of the swallowtail (Broer et al. 2007a, 2007b). For a reversible analogue see (Broer et al. 2007d, 2008).

### Remark

- The example concerns a strong resonance and it fits in some of the larger gaps of the ‘Cantor’

set described in the former application of Parametrized KAM Theory. Apart from this, the previous remarks largely apply again. It turns out that in these and many other cases there is an infinite regression regarding the resonant bifurcation diagram.

- The combination of KAM Theory with Normal Form Theory generally has been very fruitful. In the example of section “[Applications](#)” it implies that each KAM torus corresponds to a parameter value  $\lambda_0$  that is the Lebesgue density point of such quasi-periodic tori. In the real analytic case by application of (Neishtadt 1984), it can be shown that the relative measure with non-KAM tori is exponentially small in  $\lambda - \lambda_0$ .

Similar results in the real analytic Hamiltonian KAM Theory (often going by the name of exponential condensation) have been obtained by (Jorba and Villanueva 1997a, b, 2001), also compare with (Bruno 1989, 2000), ► “[Diagrammatic Methods in Classical Perturbation Theory](#)” and with (Arnold 1988; Broer and Sevryuk 2008) and references therein.

### On the Averaging Theorem

Another class of applications of normalizing-averaging is in the direction of the Averaging Theorem. There is a wealth of literature in this direction, that is not in the scope of the present paper, for further reading compare with (Arnold 1980; Arnold 1983; Arnold 1988; Sanders et al. 1985) and references therein.

**Example (A simple averaging theorem (Arnold 1980))** Given is a vector field

$$\begin{aligned} \dot{x} &= \omega(y) + \varepsilon f(x, y) \\ \dot{y} &= \varepsilon g(x, y) \end{aligned}$$

with  $(x, y) \in \mathbb{T}^1 \times \mathbb{R}^n$ , compare with (4). Roughly following the recipe of the normalization process, a suitable near-identity transformation

$$(x, y) \mapsto (x, \eta)$$

of  $\mathbb{T}^1 \times \mathbb{R}^n$  yields the following reduction, after truncating at order  $O(\varepsilon^2)$ :

$$\dot{\eta} = \varepsilon \bar{g}(\eta), \quad \text{where} \quad \bar{g}(\eta) = \frac{1}{2\pi} \int_0^{2\pi} g(x, \eta) dx.$$

We now compare  $y = y(t)$  and  $\eta = \eta(t)$  with coinciding initial values  $y(0) = \eta(0)$  as  $t$  increases. The Averaging Theorem asserts that if  $\omega(\eta) > 0$  is bounded away from 0, it follows that, for a constant  $c > 0$ , one has

$$|y(t) - \eta(t)| < c\varepsilon, \quad \text{for all } t \text{ with } 0 \leq t \leq \frac{1}{\varepsilon}.$$

This theory extends to many classes of systems, for instance to Hamiltonian systems or, in various types of systems, in the immediate vicinity of a quasi-periodic torus. Further normalizing can produce sharper estimates that are polynomial in  $\varepsilon$ , while in the analytic case this even extends over exponentially long time intervals, usually known under the name of Nekhoroshev estimates (Nekhoroshev 1971, 1977, 1979), for a description and further references also see (Broer and Sevryuk 2008), ► “Nekhoroshev Theory”. Another direction of generalization concerns passages through resonance, which in the example implies that the condition on  $\omega(\eta)$  is no longer valid. We here mention (Neishtadt 1975), for further references and descriptions referring to (Arnold 1983, 1988) and to (Sanders et al. 1985).

## Future Directions

The area of research in Normal Form Theory develops in several directions, some of which are concerned with computational aspects, including the nilpotent case. Although for smaller scale projects much can be done with computer packages, for the large scale computations, e.g., needed in Celestial Mechanics, single purpose formula manipulators have to be built. For an overview of such algorithms, compare with (Murdock 2003; Simó 1994; Simó 1998, 2000) and references therein. Here also Gevrey asymptotics is of importance.

Another direction of development is concerned with applications in Bifurcation Theory, often in particular Singularity Theory. In many of

these applications, certain coefficients in the truncated normal form are of interest and their computation is of vital importance. For an example of this in the Hamiltonian setting see (Broer and Simó 2000), where the Giorgili–Galgani (1978) algorithm was used to obtain certain coefficients at all arbitrary order in an efficient way. For other examples in this direction see (Broer et al. 2007c; Broer and Vegter 2008).

Related to this is the problem how to combine the normal form algorithms as related to the present paper, with the polynomial normal forms of Singularity Theory. The latter have a universal (i.e., context independent) geometry in the product of state space and parameter space. A problem of relevance for applications is to pull-back the Singularity Theory normal form back to the original system. For early attempts in this direction see (Broer et al. 1998a, 2003a), which, among other things, involve techniques.

## Bibliography

### Primary Literature

- Arnold VI (1980) Mathematical methods of classical mechanics. Springer, New York (English, Russian original)
- Arnold VI (1983) Geometrical methods in the theory of ordinary differential equations. Springer, New York (English, Russian original)
- Arnold VI (ed) (1988) Dynamical systems III, encyclopædia of mathematical sciences, vol 3. Springer, Berlin (English, Russian original)
- Baldomá I, Haro A (2008) One dimensional invariant manifolds of Gevrey type in real-analytic maps. DCDS series B 10(2&3):295–322
- Birkhoff GD (1927) Dynamical systems. AMS Publications
- Braaksma BLJ, Broer HW (1987) On a quasi-periodic Hopf bifurcation. Ann Inst Henri Poincaré, Analyse Non Linéaire 4(2):115–168
- Broer HW (1981) Formal normal form theorems for vector fields and some consequences for bifurcations in the volume preserving case. In: Rand DA, Young LS (eds) Dynamical systems and turbulence. Warwick, 1980 LNM 898. Springer, Berlin, pp. 54–74
- Broer HW (1993) Notes on perturbation theory 1991. In: Erasmus ICP (ed) Mathematics and fundamental applications. Aristotle University Thessaloniki, p 44
- Broer HW, Roussarie R (2001) Exponential confinement of chaos in the bifurcation set of real analytic diffeomorphisms. In: Broer HW, Krauskopf B, Vegter



- G (eds) Global analysis of dynamical systems. Festschrift dedicated to Floris Takens for his 60th birthday. Bristol Phila IOP, pp 167–210
- Broer HW, Sevryuk MB (2008) kam Theory: quasi-periodicity in dynamical systems. In: Broer HW, Hasselblatt B, Takens F (eds) Handbook of Dynamical Systems, vol 3. North-Holland (to appear 2009)
- Broer HW, Simó C (2000) Resonance tongues in Hill's equations: a geometric approach. *J Differ Equ* 166: 290–327
- Broer HW, Takens F (1989) Formally symmetric normal forms and genericity. *Dyn Rep* 2:36–60
- Broer HW, Tangerman FM (1986) From a differentiable to a real analytic perturbation theory, applications to the Kupka Smale theorems. *Ergod Theor Dyn Syst* 6: 345–362
- Broer HW, Vegter G (1992) Bifurcational aspects of parametric resonance. *Dyn Rep New Ser* 1:1–51
- Broer HW, Vegter G (2008) Generic Hopf-Neimark-Sacker bifurcations in feed forward systems. *Nonlinearity* 21: 1547–1578
- Broer HW, Huitema GB, Takens F, Braaksma BLJ (1990) Unfoldings and bifurcations of quasi-periodic tori. *Mem AMS* 83(421):i–vii; pp 1–175
- Broer HW, Dumortier F, van Strien SJ, Takens F (1991) Structures in dynamics: finite dimensional deterministic studies. In: van Groesen EWC, de Jager EM (eds) Studies in mathematical physics, vol 2. North-Holland, Amsterdam (Russian translation 2003)
- Broer HW, Chow SN, Kim Y, Vegter G (1993) A normally elliptic Hamiltonian bifurcation. *ZAMP* 44:389–432
- Broer HW, Huitema GB, Sevryuk MB (1996a) Quasi-periodic motions in families of dynamical systems: order amidst chaos. *LNM 1645* Springer, Berlin
- Broer HW, Roussarie R, Simó C (1996b) Invariant circles in the Bogdanov-Takens bifurcation for diffeomorphisms. *Ergod Theor Dyn Syst* 16:1147–1172
- Broer HW, Hoveijn I, Lunter GA, Vegter G (1998a) Resonances in a Spring–Pendulum: algorithms for equivariant singularity theory. *Nonlinearity* 11(5):1–37
- Broer HW, Simó C, Tatjer JC (1998b) Towards global models near homoclinic tangencies of dissipative diffeomorphisms. *Nonlinearity* 11:667–770
- Broer HW, Takens F, Wagener FOO (1999) Integrable and non-integrable deformations of the skew Hopf bifurcation. *Reg Chaotic Dyn* 4(2):17–43
- Broer HW, Hoveijn I, Lunter GA, Vegter G (2003a) Bifurcations in Hamiltonian systems: computing singularities by Gröbner bases. *LNM*, vol 1806. Springer, Berlin
- Broer HW, Golubitsky M, Vegter G (2003b) The geometry of resonance tongues: a singularity theory approach. *Nonlinearity* 16:1511–1538
- Broer HW, Hanßmann H, Jorba A, Villanueva J, Wagener FOO (2003c) Normal-internal resonances in quasi-periodically forced oscillators: a conservative approach. *Nonlinearity* 16:1751–1791
- Broer HW, Hanßmann H, Hoo J (2007a) The quasi-periodic Hamiltonian Hopf bifurcations. *Nonlinearity* 20:417–460
- Broer HW, Hoo J, Naudot V (2007b) Normal linear stability of quasi-periodic tori. *J Differ Equ* 232(2):355–418
- Broer HW, Golubitsky M, Vegter G (2007c) Geometry of resonance tongues. In: Chéniet D, Dutertre N, Murolo C, Trotman D, Pichon A (eds) Proceedings of the 2005 marseille singularity school and conference, dedicated to Jean-Paul Brasselet on his 60th birthday. World Scient, pp. 327–356
- Broer HW, Ciocci MC, Hanßmann H (2007d) The quasi-periodic reversible Hopf bifurcation. *IJBC* 17(8): 2605–2623
- Broer HW, Naudot V, Roussarie R, Saleh K, Wagener FOO (2007e) Organising centres in the semi-global analysis of dynamical systems. *IJAMAS* 12(D07):7–36
- Broer HW, Ciocci MC, Hanßmann H, Vanderbauwhede A (2008) Quasi-periodically stable unfoldings of normally resonant tori. *Physica D* (to appear)
- Broer HW, Hanßmann H, You J () On the destruction of resonant Lagrangean tori in Hamiltonian systems. (in preparation)
- Bruno AD (1965) On the convergence of transformations of differential equations to the normal form. *Soviet Math Dokl* 6:1536–1538 (English, Russian original)
- Bruno AD (1989) Local methods in nonlinear differential equations. Springer, Berlin (English, Russian original)
- Bruno AD (2000) Power geometry in algebraic and differential equations. North-Holland Mathematical Library, vol 57. North-Holland, Amsterdam (English, Russian original)
- Chow SN, Li C, Wang D (1994) Normal forms and bifurcations of planar vector fields. Cambridge University Press, Cambridge
- Ciocci MC, Litvak-Hinzenon A, Broer HW (2005) Survey on dissipative KAM theory including quasi-periodic bifurcation theory based on lectures by Henk Broer. In: Montaldi J, Ratiu T (eds) Geometric mechanics and symmetry: the Peyresq lectures. LMS lecture notes series, vol 306. Cambridge University Press, Cambridge, pp 303–355
- Cushman RH, Sanders JA (1986) Nilpotent normal forms and representation theory of  $sl(2, \mathbb{R})$ . In: Golubitsky M, Guckenheimer J (eds) Multi-parameter bifurcation theory. AMS. Providence, pp. 31–51
- Deprit A (1969) Canonical transformations depending on a small parameter. *Celest Mech* 1:12–30
- Deprit A (1981) The elimination of the parallax in satellite theory. *Celest Mech* 24:111–153
- Eliasson LH, Kuksin SB, Marmi S, Yoccoz JC (2002) Dynamical systems and small divisors. *LNM*, vol 1784. Springer, Berlin
- Ferrer S, Hanßmann H, Palacián J, Yanguas P (2002) On perturbed oscillators in 1-1-1 resonance: the case of axially symmetric cubic potentials. *J Geom Phys* 40: 320–369
- Galin DM (1982) Versal deformations of linear Hamiltonian systems. *Am Soc Trans Ser* 2(118):1–12
- Giorgilli A, Galgani L (1978) Formal integrals for an autonomous Hamiltonian system near an equilibrium point. *Celest Mech* 17:267–280



- Giorgilli A, Delshams A, Fontich E, Galgani L, Simó C (1989) Effective stability for a Hamiltonian system near an elliptic equilibrium point with an application to the restricted three body problem. *J Differ Equ* 77: 167–198
- Gustavson FG (1966) On constructing formal integrals of a Hamiltonian system near an equilibrium point. *Astron J* 71:670–686
- Hanßmann H (2007) Local and semi-local bifurcations in hamiltonian dynamical systems. LNM, vol 1893. Springer, Berlin
- Hoveijn I (1996) Versal deformations and normal forms for reversible and Hamiltonian linear systems. *J Diff Eqns* 126(2):408–442
- Iooss G (1988) Global characterization of the normal form of a vector field near a closed orbit. *J Diff Eqns* 76: 47–76
- Jorba À, Villanueva J (1997a) On the persistence of lower dimensional invariant tori under quasi-periodic perturbations. *J Nonlinear Sci* 7:427–473
- Jorba À, Villanueva J (1997b) On the normal behaviour of partially elliptic lower-dimensional tori of Hamiltonian systems. *Nonlinearity* 10:783–822
- Jorba À, Villanueva J (2001) The fine geometry of the Cantor families of invariant tori in Hamiltonian systems. In: Casacuberta C, Miró-Roig RM, Verdera J, Xambó-Descamps S (eds) *European Congress of Mathematics, Barcelona, 2000, vol 2. Progress in Mathematics*, vol 202. Birkhäuser, Basel, pp 557–564
- Koçak H (1984) Normal forms and versal deformation of linear Hamiltonian systems. *J Diff Eqns* 51:359–407
- Marco JP, Sauzin D (2002) Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems. *Publ IHES* 96:199–275
- Marsden JE, McCracken M (1976) *The Hopf-bifurcation and its applications*. Springer, New York
- Meyer KR (1984) Normal forms for the general equilibrium. *Funkcial Ekva* 27:261–271
- Milnor JW (2006) *Dynamics in one complex variable*, 3rd edn. *Ann Math Stud*, vol 160. Princeton University Press, Princeton
- Moser JK (1968) Lectures on Hamiltonian systems. *Mem Am Math Soc* 81:1–60
- Murdock J (2003) *Normal forms and unfoldings for local dynamical systems*. Springer monographs in mathematics. Springer, New York
- Narasimhan R (1968) *Analysis on real and complex manifolds*. Elsevier, North Holland
- Neishtadt AI (1975) Passage through resonances in two-frequency problem. *Sov Phys Dokl* 20(3):189–191 (English, Russian original)
- Neishtadt AI (1984) The separation of motions in systems with rapidly rotating phase. *J Appl Math Mech* 48(2): 133–139 (English, Russian original)
- Nekhoroshev NN (1971) On the behavior of Hamiltonian systems close to integrable ones. *Funct Anal Appl* 5: 338–339 (English, Russian original)
- Nekhoroshev NN (1977) An exponential estimate of the stability time of nearly integrable Hamiltonian systems. I, *Russian Math Surv* 32(6):1–65 (English, Russian original)
- Nekhoroshev NN (1979) An exponential estimate of the stability time of nearly integrable Hamiltonian systems, vol II. *Trudy Sem Imeni. In: Petrovskogo IG* pp 5:5–50 (in Russian) English translation. In: Oleinik OA (ed) (1985) *Topics in modern mathematics, Petrovskii Seminar. Consultants Bureau, New York*, vol 5, pp 1–58
- Palis J, de Melo WC (1982) *Geometric theory of dynamical systems*. Springer
- Palis J, Takens F (1993) *Hyperbolicity & sensitive chaotic dynamics at homoclinic bifurcations*. Cambridge studies in advanced mathematics, Cambridge University Press 35
- Poincaré H (1928) *Euvres*, vol I. Gauthier-Villars
- Popov G (2000) Invariant tori, effective stability, and quasimodes with exponentially small error terms. I: Birkhoff normal forms. *Ann Henri Poincaré* 1(2): 223–248
- Popov G (2004) KAM theorem for Gevrey Hamiltonians. *Ergod Theor Dynam Syst* 24:1753–1786
- Ramis JP (1994) Séries divergentes et théories asymptotiques. *Panor Synth* pp 0–74
- Ramis JP, Schäfke R (1996) Gevrey separation of slow and fast variables. *Nonlinearity* 9:353–384
- Roussarie R (1987) Weak and continuous equivalences for families of line diffeomorphisms. In: *Dynamical systems and bifurcation theory, Pitman research notes in math. Series Longman* 160:377–385
- Sanders JA, Verhulst F, Murdock J (1985) *Averaging methods in nonlinear dynamical systems*. Revised 2nd edn, *Appl Math Sciences* 59, 2007. Springer
- Siegel CL (1942) Iteration of analytic functions. *Ann Math* 43(2):607–612
- Siegel CL, Moser JK (1971) *Lectures on celestial mechanics*. Springer, Berlin
- Simó C (1994) Averaging under fast quasiperiodic forcing. In: Seimenis J (ed) *Hamiltonian mechanics, integrability and chaotic behaviour*. NATO Adv Sci Inst Ser B Phys 331:13–34, Plenum, New York
- Simó C (1998) Effective computations in celestial mechanics and astrodynamics. In: Rumyantsev VV, Karapetyan AV (eds) *Modern methods of analytical mechanics and their applications*. CISM Courses Lectures, vol 387. Springer, pp 55–102
- Simó C (2000) Analytic and numeric computations of exponentially small phenomena. In: Fiedler B, Gröger K, Sprekels J (eds) *Proceedings EQUADIFF 99*, Berlin. World Scientific, Singapore, pp 967–976
- Spivak M (1970) *Differential Geometry*, vol I. Publish or Perish Inc
- Sternberg S (1959) On the structure of local homeomorphisms of Euclidean  $n$ -space, vol II. *Am J Math* 81: 578–605
- Takens F (1973) *Forced oscillations and bifurcations*. Applications of global analysis, vol I, Utrecht. Comm Math Inst Univ Utrecht, pp 31–59 (1974). Reprinted In: Broer HW, Krauskopf B, Vegter G (eds) (2001) *Global*

- analysis of dynamical systems. Festschrift dedicated to Floris Takens for his 60th birthday Leiden. Inst Phys Bristol, pp 1–61
- Takens F (1974) Singularities of vector fields. Publ Math IHES 43:47–100
- Takens F, Vanderbauwhede A (2009) Local invariant manifolds and normal forms. In: Broer HW, Hasselblatt B, Takens F (eds) Handbook of dynamical systems, vol 3. North-Holland (to appear)
- Thom R (1989) Structural stability and morphogenesis. An outline of a general theory of models, 2nd edn. Addison-Wesley, Redwood City. (English, French original)
- van der Meer JC (1985) The hamiltonian Hopf bifurcation. LNM, vol 1160. Springer
- Vanderbauwhede A (1989) Centre manifolds, normal forms and elementary bifurcations. Dyn Rep 2:89–170
- Varadarajan VS (1974) Lie groups, Lie algebras and their representations. Prentice-Hall, Englewood Cliffs
- Wagener FOO (2003) A note on Gevrey regular KAM theory and the inverse approximation lemma. Dyn Syst 18(2):159–163
- Wagener FOO (2005) On the quasi-periodic  $d$ -fold degenerate bifurcation. J Differ Eqn 216:261–281
- Yoccoz JC (1995) Théorème de Siegel, nombres de Bruno et polynômes quadratiques. Astérisque 231:3–88

## Books and Reviews

- Braaksma BLJ, Stolovitch L (2007) Small divisors and large multipliers (Petits diviseurs et grands multiplieurs). Ann l'institut Fourier 57(2):603–628
- Broer HW, Levi M (1995) Geometrical aspects of stability theory for Hill's equations. Arch Rat Mech An 131: 225–240
- Gaeta G (1999) Poincaré renormalized forms. Ann Inst Henri Poincaré 70(6):461–514
- Martinet J, Ramis JP (1982) Problèmes des modules pour les équations différentielles non linéaires du premier ordre. Publ IHES 55:63–164
- Martinet J, Ramis JP (1983) Classification analytique des équations différentielles non linéaires résonnantes du premier ordre. Ann Sci École Norm Supérieure Sér 4 16(4):571–621
- Vanderbauwhede A (2000) Subharmonic bifurcation at multiple resonances. In: Elaydi S, Allen F, Elkhader A, Mughrabi T, Saleh M (eds) Proceedings of the mathematics conference, Birzeit, August 1998. World Scientific, Singapore, pp 254–276

# Convergence of Perturbative Expansions

Sebastian Walcher  
Lehrstuhl A für Mathematik, RWTH Aachen,  
Aachen, Germany

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Poincaré–Dulac Normal Forms  
Convergence and Convergence Problems  
Lie Algebra Arguments  
NFIM and Sets of Analyticity  
Hamiltonian Systems  
Future Directions  
Bibliography

## Glossary

**Lie algebras of vector fields** The vector space of analytic vector fields on an open subset  $U$  of  $\mathbb{C}^n$ , with the bracket  $[p, q]$  defined by

$$[p, q](x) = Dq(x)p(x) - Dp(x)q(x)$$

becomes a Lie algebra, as is well known. Mutatis mutandis, this also holds for local analytic and for formal vector fields. As noted previously, PDNF is most naturally defined via this Lie bracket. Moreover, the “algebraic” obstructions to convergence are most appropriately discussed within the Lie algebra framework.

**Normal form on invariant manifolds** While there may not exist a convergent transformation to PDNF for a given vector field  $f$ , one may have a convergent transformation to a “partially normalized” vector field, which admits a certain invariant manifold and is in PDNF when restricted to this manifold. This observation is of some practical importance.

## Normalizing transformations and convergence

A relatively straightforward argument shows that any formal vector field  $f = B + \dots$  can be transformed to a formal vector field in PDNF via a formal power series transformation. But for analytic vector fields, the existence of a convergent transformation is not assured. There are two obstacles to convergence: First, the possible existence of small denominators (roughly, this means that the eigenvalues satisfy “near-resonance conditions”); and second, “algebraic” obstructions due to the particular form of the normalized vector field.

**Poincaré–Dulac normal form** Let  $f = B + \dots$  be a formal or analytic vector field about 0, and let  $B_s$  be the semisimple part of  $B$ . Then one says that  $f$  is in Poincaré–Dulac normal form (PDNF) if  $[B_s, f] = 0$ . An equivalent characterization, if  $B$  is in Jordan form, is to say that only resonant monomials occur in the series expansion.

**Resonant eigenvalues** Let  $B$  be a linear endomorphism of  $\mathbb{C}^n$ , with eigenvalues  $\lambda_1, \dots, \lambda_n$  (counted according to multiplicity). One calls these eigenvalues *resonant* if there are integers  $d_j \geq 0$ ,  $\sum d_j \geq 2$ , and some  $k \in \{1, \dots, n\}$  such that

$$d_1\lambda_1 + \dots + d_n\lambda_n - \lambda_k = 0.$$

If  $B$  is represented by a matrix in Jordan canonical then the associated vector monomial  $x_1^{d_1} \dots x_n^{d_n} e_k$  will be called a *resonant monomial*. (Here  $e_1, \dots, e_n$  denote the standard basis, and the  $x_i$  are the corresponding coordinates.)

## Definition of the Subject

It seems appropriate to first clarify what types of perturbative expansions are to be considered here. There exist various types of such expansions in various settings (a very readable introduction is given in Verhulst’s monograph (2005)), but for many of these settings the question of

convergence is not appropriate or irrelevant. Therefore we restrict attention to the scenario outlined, for instance, in the introductory chapter of the monograph (1999) by Cicogna and Gaeta, which means consideration of normal forms and normalizing transformation for local analytic-vector fields.

Normal forms are among the most important tools for the local analysis and classification of vector fields and maps near a stationary point (see ► “Normal Forms in Perturbation Theory”). Convergence problems arise here, and they turn out to be surprisingly complex. While the first contributions date back more than a century, some very strong and very deep results are just a few years old, and this remains an active area of research. Clearly convergence questions are relevant for the analytic classification of local vector fields, but they are also of practical relevance in applications, e.g., for stability questions and for the existence of particular types of solutions.

## Introduction

The theory of normal forms was initiated by Poincaré (1879), and later extended by Dulac (1912), and by Birkhoff (1927) to Hamiltonian vector fields. There exist various types of normal forms, depending on the specific problem one wants to address. Bruno (1971, 1989) in the 1960s and 1970s performed a comprehensive and deep investigation of Poincaré–Dulac normal forms, which are defined with respect to the semi-simple part of the linearization. Such normal forms are very important in applications, and moreover they have certain built-in symmetries, which allows a well-defined reduction procedure.

We will first give a quick review of normalization procedures and normal forms, and then discuss convergence problems (which mostly refers to convergence or divergence of normalizing transformations). We will present fundamental convergence and divergence results due to Poincaré, Siegel, Bruno and others. We then proceed to discuss the relevance of certain Lie algebras of analytic vector fields for these matters, including results by Cicogna and the author of

this article on the influence of symmetries, and the far-reaching generalization of Bruno’s theorems (among others) due to Stolovitch. Variants of normal forms which guarantee convergence on certain subsets (due to Bibikov and Bruno) are then discussed, and applications are mentioned. Finally, the Hamiltonian setting, which deserves a discussion in its own right, is presented, starting with results of Ito and recently culminating in Zung’s convergence theorem. For Hamiltonian systems there is also work due to Perez–Marco on divergence of normal forms.

Within the space limitations of this contribution, and in view of some very intricate and space-consuming technical questions and conditions, the author tried to find an approach that, for some problems, should provide some insight into a result, the arguments in its proof or its relevance, without exhibiting all the technicalities, or without giving the most general statement. The author hopes to have been somewhat successful in this, and apologizes to the creators of the original theorems for presenting just “light” versions.

## Poincaré–Dulac Normal Forms

We will start with a coordinate-free approach to normal form theory. Our objects are local ordinary differential equations (over  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ )

$$\dot{x} = F(x), \quad F(0) = 0$$

with  $F$  analytic, thus we have a convergent series expansion

$$F(x) = Bx + \sum_{j \geq 2} f_j(x) = Bx + f_2(x) + f_3(x) + \cdots \quad \text{near } 0 \in \mathbb{K}^n.$$

Here  $B = DF(0)$  is linear, and each  $f_j$  is homogeneous of degree  $j$ . Our objective is to simplify the Taylor expansion of  $F$ . For this purpose, take an analytic “near-identity” map

$$H(x) = x + h_2(x) + \cdots$$

Since  $H$  is locally invertible, there is a unique

$$F^*(x) = Bx + \sum_{j \geq 2} f_j^*(x)$$

such that the identity

$$DH(x)F^*(x) = F(H(x)) \quad (\text{R})$$

holds.  $H$  “preserves solutions” in the sense that parametrized solutions of  $\dot{x} = F^*(x)$  are mapped to parametrized solutions of  $\dot{x} = F(x)$  by  $H$ . It is convenient to introduce the following abbreviation:

$$F^* \xrightarrow{H} F \quad \text{if } (\text{R}) \text{ holds.}$$

Given the expansion

$$F(x) = Bx + f_2(x) + \dots + f_{r-1}(x) + f_r(x) + \dots,$$

assume that  $f_2, \dots, f_{r-1}$  are already deemed “satisfactory” (according to some specified criterion). Then the ansatz  $H(x) = x + h_r(x) + \dots$  yields

$$F^*(x) = Bx + f_2(x) + \dots + f_{r-1}(x) + f_r^*(x) + \dots$$

(with terms of degree  $< r$  unchanged), and at degree  $r$  one obtains the so-called *homological equation*

$$[B, h_r] = f_r - f_r^*$$

(Here  $[p, q](x) = Dq(x)p(x) - Dp(x)q(x)$  denotes the usual Lie bracket of vector fields).

The space  $\mathcal{P}_r$  of homogeneous vector polynomials of degree  $r$  is finite dimensional, and  $\text{ad } B = [B, \cdot]$  sends  $\mathcal{P}_r$  to  $\mathcal{P}_r$ . Thus the homological equation poses a linear algebra problem on a finite dimensional vector space: Given  $B$  and  $f_r$ , determine  $f_r^*$  so that the equation can be solved and let  $h_r$  be a solution. How can  $f_r^*$  be chosen? Let  $\mathcal{W}$  be any subspace of  $\mathcal{P}_r$  such that

$$\text{image}(\text{ad } B) + \mathcal{W} = \mathcal{P}_r.$$

Then one may choose  $f_r^* \in \mathcal{W}$ . If the sum is direct then  $f_r^* \in \mathcal{W}$  is uniquely determined by  $f_r$ .

Generally, the type of normal form is specified by the choice of a subspace  $\mathcal{W}_r$  for each degree  $r$  such that  $\text{image}(\text{ad } B) + \mathcal{W}_r = \mathcal{P}_r$ . The Poincaré–Dulac choice is as follows: Given the decomposition

$$B = B_s + B_n$$

into semisimple and nilpotent part, then  $\text{ad } B = \text{ad } B_s + \text{ad } B_n$  is known to be the corresponding decomposition on  $\mathcal{P}_r$ . Choose  $\mathcal{W}_r = \text{Ker}(\text{ad } B_s)$ . By linear algebra

$$\mathcal{W}_r + \text{image}(\text{ad } B) = \mathcal{P}_r;$$

and the sum is direct if  $B$  is semisimple. In any case we have  $[B_s, f_r^*] = 0$ .

**Definition 1** The vector field  $F^*$  is in *Poincaré–Dulac normal form* if  $[B_s, f_j^*] = 0$  for all  $j$ ; equivalently if  $[B_s, F^*] = 0$ .

If one sets aside convergence questions for the moment, an immediate consequence of the considerations above is:

**Proposition 1** For any  $F = B + \dots$  there are formal power series  $H(x) = x + \dots$ ,  $F^*(x) = Bx + \dots$  such that  $F^* \xrightarrow{H} F$  and  $F^*$  is in *Poincaré–Dulac normal form*.

One should note that the homological equation is of relevance beyond the setting of Poincaré–Dulac, and forms the foundation for various other (or more refined) types of normal form. See also the entry in this section of the Encyclopedia by T. Gramchev on normal forms with respect to a nilpotent linear part.

Now let us turn to the standard approach via suitable coordinates. Given  $F$  be as above, let  $\lambda_1, \dots, \lambda_n$  be the eigenvalues of  $B$ . Complexify, if necessary, and assume that  $B$  is in Jordan canonical form with respect to the given coordinates  $x_1, \dots, x_n$  (with corresponding basis  $e_1, \dots, e_n$  of  $\mathbb{K}^n$ ). In particular,

$$B_s = \text{diag}(\lambda_1, \dots, \lambda_n).$$

In the following we will use  $x_i$  and  $e_i$  to denote eigencoordinates, resp. eigenbasis elements, and

reserve  $\lambda_1, \dots, \lambda_n$  for the eigenvalues of  $B$ , without explicitly saying so in every instance.

**Lemma 1** The “vector monomial”  $p(x) = x_1^{m_1} \dots x_n^{m_n} e_j$  satisfies

$$[B_s, p] = (m_1 \lambda_1 + \dots + m_n \lambda_n - \lambda_j) \cdot p$$

Thus, the vector monomials form an eigenbasis of  $\text{ad } B_s$  on the space  $\mathcal{P}_r$ , with eigenvalues  $m_1 \lambda_1 + \dots + m_n \lambda_n - \lambda_j$  ( $m_j$  nonnegative integers,  $\sum m_j = r$ ).

The eigenvalues play a crucial role both for the classification of formal normal forms and for convergence issues. The following distinction is pertinent here:

**Definition 2** One calls  $(\lambda_1, \dots, \lambda_n)$  *resonant* if there are integers  $d_j \geq 0$ ,  $\sum d_j \geq 2$ , and some  $k \in \{1, \dots, n\}$  such that

$$d_1 \lambda_1 + \dots + d_n \lambda_n - \lambda_k = 0.$$

In this case, the vector monomial  $x_1^{d_1} \dots x_n^{d_n} e_k$  is also called a *resonant monomial*. One calls  $(\lambda_1, \dots, \lambda_n)$  *non-resonant* otherwise.

Given eigencoordinates, one can characterize a Poincaré–Dulac normal form by the property that only resonant monomials occur in the series expansion. Moreover, evaluation of the homological equation shows that the nonzero eigenvalues of  $\text{ad } B$  will occur as denominators in its solution. To iterate the normalization, one may proceed degree by degree with a series of transformations of the form  $\exp(h_r)$ , using the solution of the homological equation; see Walcher (1993). One may also choose a different iterative approach to compute a normalizing transformation, such as the important “distinguished transformation” of Bruno (1971). In any case this will yield coefficients whose denominators are products of nonzero terms  $m_1 \lambda_1 + \dots + m_n \lambda_n - \lambda_j$ . This is the source of convergence problems caused by *small denominators*:

There are formal series  $H$  and  $F^*$  such that  $F^* \xrightarrow{H} F$  and  $F^*$  is in Poincaré–Dulac normal form, but does there exist a convergent  $H$ ? (Here

“convergent” means: convergent in some neighborhood of 0.)

To answer this question, it is necessary to specify a particular type of transformation: Transformations to Poincaré–Dulac normal form are not necessarily unique, even if one stipulates a near-identity transformation  $H(x) = x + \dots$ . Non-uniqueness occurs whenever the eigenvalues of  $B$  are resonant, because then some homological equation will not have a unique solution: In eigencoordinates of  $B_s$ , the series of the transformation is fixed only up to resonant monomial terms. In particular, if  $F$  itself is in normal form then any formal power series  $H(x) = x + \dots$  that contains only resonant monomials will provide a normal form  $F^*$ , and a suitable nontrivial choice of  $H$  will even force  $F^* = F$ . Thus there may be divergent transformations sending an equation in normal form to itself.

## Convergence and Convergence Problems

Let us note at the start that for given analytic  $F$  there may not exist any convergent transformation to normal form; thus the convergence question has no simple answer.

An early positive convergence result is due to Poincaré (1879), with a later improvement due to Dulac (1912):

**Theorem 1 (Poincaré–Dulac)** *If  $\lambda_1, \dots, \lambda_n$  lie in an open half-plane in  $\mathbb{C}$  which does not contain 0 then there exists a convergent transformation.*

For example, one may think of the open left half-plane. The proof is relatively straightforward, employing natural majorants. (Due to the hypothesis, the  $|\sum_{i=1}^n m_i \lambda_i - \lambda_j|$  are unbounded for  $\sum m_i \rightarrow \infty$ .) Poincaré’s condition does not preclude the existence of resonant monomials but it ensures that there are at most finitely many of these.

One main technical difficulty in proving convergence was to replace the direct majorant arguments by more efficient, and more sophisticated, tools, so that small denominator problems could be tackled. The following result, due to C. L.

Siegel (1952), may be seen as the start of the “modern phase” for convergence results. Characteristically, this result goes back to a mathematician who also worked in analytic number theory. Siegel assumed that the eigenvalues satisfy a certain arithmetic condition.

**Condition S:** The eigenvalues are pairwise different and there are constants  $C > 0$ ,  $\nu > 0$  such that for all nonnegative integer tuples  $(m_i)$ ,  $\sum m_i > 1$ , the following inequality holds:

$$\left| \sum_{i=1}^n m_i \lambda_i - \lambda_j \right| \geq C \cdot (m_1 + \dots + m_n)^{-\nu}$$

**Theorem 2 (Siegel)** *If Condition S holds then there is a convergent transformation to normal form.*

**Proof** A very rough sketch of the proof is as follows. One works in eigencoordinates. By scaling one may assume that all coefficients in the expansion of  $F$  are absolutely bounded by some constant  $M \geq 1$ . For the transformation one writes

$$H(x) = x + \sum_{r \geq 2} \left( \sum \alpha_{m_1, \dots, m_n, k} \cdot x_1^{m_1} \cdots x_n^{m_n} e_k \right)$$

where the sum inside the bracket extends over all nonnegative integer tuples with  $\sum m_i = r$  and  $1 \leq k \leq n$ . (Since Condition S precludes resonances, the coefficients of the series are uniquely determined.) Now set

$$A_{m_1, \dots, m_n} := \sum_k |\alpha_{m_1, \dots, m_n, k}|.$$

From the homological equations one finds by recursion

$$\left| \sum_i m_i \lambda_i - \lambda_k \right| \cdot |\alpha_{m_1, \dots, m_n, k}| \leq M \cdot \sum A_{d_{1,1}, \dots, d_{1,n}} \cdots A_{d_{s,1}, \dots, d_{s,n}}$$

where the summation on the right hand side extends over all tuples  $(d_{i,1}, \dots, d_{i,n})$  that add up to  $(m_1, \dots, m_n)$ . From this one may obtain an

estimate for  $A_{m_1, \dots, m_n}$ , and invoking Condition S one eventually arrives at the conclusion that  $\sum A_{m_1, \dots, m_n} x_1^{m_1} \cdots x_n^{m_n}$  is majorized by the series of  $\frac{x_1 + \dots + x_n}{1 - K(x_1 + \dots + x_n)}$ , some  $K > 0$ .  $\square$

**Example** Let  $\dot{x} = Bx + \dots$  be given in dimension two, and assume that the eigenvalues  $\lambda_1, \lambda_2$  of  $B$  are nonresonant, and are algebraic irrational numbers. (This is the case when the entries of  $B$  are rational but the characteristic polynomial is irreducible over the rationals.) Then  $\lambda_2/\lambda_1$  is algebraic but not rational, and  $(\lambda_1, \lambda_2)$  satisfies Condition S, due to a celebrated number-theoretic result of Thue, Siegel and Roth. Thus there exists a convergent transformation to normal form.

While Siegel’s convergence proof uses majorizing series, the approach is not as straightforward as in the Poincaré setting. Siegel’s result is strong in the sense that Condition S is satisfied by Lebesgue – almost all tuples  $(\lambda_1, \dots, \lambda_n) \in \mathbb{C}^n$ . But the condition forces the normal form to be uninteresting: One necessarily has  $F^* = B = B_s$ . In the same paper (1952), Siegel also notes that divergence is possible, even for a set of “eigenvalue vectors” that is everywhere dense in  $n$ -space.

In the resonant case there is a second source of obstacles to convergence. An early example for this is due to Horn (about 1890); see Bruno (1971):

**Example** The system

$$\begin{aligned} \dot{x}_1 &= x_1^2 \\ \dot{x}_2 &= x_2 - x_1 \end{aligned}$$

(with eigenvalues  $(0, 1)$  for the linear part) admits no convergent transformation to normal form. A detailed proof for this can be found in Cicogna and Walcher (2002). The underlying reason is that the ansatz for a transformation – unavoidably – leads to the differential equation

$$x^2 \cdot y' = y - x,$$

(which goes back to Euler) with divergent solution  $\sum_{k \geq 1} (k-1)! x^k$ . There are no small denominators here. The problem actually lies within the normal form, which can be computed as



$$\begin{aligned}\dot{y}_1 &= y_1^2 \\ \dot{y}_2 &= y_2.\end{aligned}$$

The single nonlinear term is sufficient to obstruct convergence.

Pliss (1965) showed that Siegel's theorem still holds if there are no such nonlinear obstructions in the normal form:

**Theorem 3 (Pliss)** *Assume that:*

- (i) *The nonzero elements among the  $\sum_{i=1}^n m_i \lambda_i - \lambda_j$  satisfy Condition S.*
- (ii) *Some formal normal form of  $F$  is equal to  $B = B_s$ .*

*Then there exists a convergent transformation to normal form.*

While it seemingly extends Siegel's result only to a rather narrow special setting, Pliss' theorem proved to be quite important for future developments. Pliss uses a different approach to proving convergence, via a generalized Newton method.

Fundamental insights into normal forms, and in particular into convergence and divergence problems were achieved by Bruno, starting in the 1960s; see (1971, 1989). His results included or surpassed much of the earlier work. Let us take a closer look at Bruno's conditions. As above, let  $B$  be in Jordan form, with eigenvalues  $\lambda_1, \dots, \lambda_n$ . For  $k \geq 1$  set

$$\omega_k := \min \left\{ \left| \sum m_i \lambda_i - \lambda_j \right| \neq 0 : 1 \leq j \leq n, \right. \\ \left. m_i \in \mathbb{Z}_+, \sum m_i < 2^k \right\}$$

Bruno introduced two arithmetic conditions: Condition  $\omega$ :

$$-\sum_{k=1}^{\infty} \frac{\ln \omega_k}{2^k} < \infty$$

Condition  $\bar{\omega}$ :

$$\limsup_k -\frac{\ln \omega_k}{2^k} < \infty$$

Condition  $\omega$  can be shown to imply Condition S. Clearly Condition  $\omega$  implies Condition  $\bar{\omega}$ .

Possibly in view of Pliss' theorem, Bruno introduced the following algebraic condition on the normal form:

*Condition A:* Some formal normal form is of the type

$$F^* = \sigma \cdot B_s$$

with  $\sigma$  a scalar formal power series.

Pliss' condition corresponds to Condition A with  $\sigma = 1$ . Moreover, one can show that Condition A (as well as Pliss' condition) is satisfied by every (formal) normalform if it is satisfied by one.

Now let us turn to Bruno's main theorems. The convergence theorem here may be expected in view of the previous results, but the divergence theorem – as well as its proof – conquers new ground. (We do not state the most general version of the divergence theorem.)

**Theorem 4 (Bruno's convergence theorem)** *If Condition  $\omega$  and Condition A are satisfied then a convergent normal form transformation exists.*

**Theorem 5 (Bruno's divergence theorem)**

*Assume that  $\lambda_1, \dots, \lambda_n$  do not lie in a complex half-plane with 0 in its boundary (in particular they do not satisfy the Poincaré condition). Moreover assume that an analytic vector field  $F^*$  in normal form does not satisfy a weaker version of Condition A, or that Condition  $\bar{\omega}$  is not satisfied. Then there exists an analytic  $F$  with normal form  $F^*$  such that no transformation of  $F$  to (any) normal form converges.*

As for the “weaker version of Condition A”, see Remark (c) below. Bruno also discusses the scenario when all eigenvalues are contained in some complex half-plane with 0 in its boundary.

**Example** There exists an analytic vector field

$$F(x) = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix} + \dots$$

which admits no convergent transformation to normal form (which is just the linear part). The arithmetic conditions on the eigenvalues are satisfied, but the nontrivial Jordan block violates Condition A. For the related problem of normalizing local analytic diffeomorphisms (rather than vector fields), see e.g., DeLatte and Gramchev (2002) and Gramchev (2002), for similar divergence results.

**Remark (a)** Bruno's divergence theorem cannot be applied directly to prove divergence of all normalizing transformations for a specific given vector field. (Therefore any particular example still has to be worked "by hand".) Rather, the divergence theorem provides a generic result. Since the arithmetic conditions (both  $\omega$  and  $\bar{\omega}$ ) are very weak, one sees that in absence of the Poincaré condition, the algebraic obstructions from the formal normal form are mostly responsible for divergence.

**Remark (b)** Bruno's divergence theorem starts from the assumption that a convergent normal form exists for a given vector field; thus he deals with convergence or divergence of normalizing transformations. Little seems to be known about analytic vector fields that admit only divergent normal forms.

**Remark (c)** The version of Condition A stated here is taken from Bruno's monograph (see Chap. III, § in Bruno 1989). The original paper (p. 140 ff. in Bruno 1971). Contains somewhat different versions. In the interesting case when the hypothesis of Theorem 5 holds, Bruno's original requirement on the normal form in eigen-coordinates is as follows:

$$\dot{x}_j = x_j(\lambda_j \sigma + \bar{\lambda}_j \tau), \quad 1 \leq j \leq n,$$

with scalar series  $\sigma$  and  $\tau$ . This condition seems to have been stated with special regard to real systems, because otherwise complex conjugation plays no distinguished role. To illustrate this point, consider the complexification of a real system, and multiply this through by some scalar exp

( $i\theta$ ) which is neither real nor purely imaginary. Convergence or divergence of normalizing transformations is not affected by this, but the shape of the condition above changes considerably. (The appropriate setting for this seems to be the one developed by Stolovitch (2000); see below.) In later publications, notably in Bruno (1989), Bruno himself mostly used Condition A in the simple form we stated above.

## Lie Algebra Arguments

Bruno's theorems set the standard against which later results are measured. However, they do not directly address this basic question: Given a particular local analytic vector field  $F$ , characterize properties that are necessary (and, ideally, also sufficient) for the existence of a convergent normal form transformation. (Obviously, Condition A is not generally necessary for the existence of a convergent normalizing transformation.) To answer this basic question and related problems, it turns out helpful to consider Lie algebras of analytic vector fields.

A possible approach is based on the earlier observation that normal forms admit symmetries. (See also the entry on ► "Symmetry and Perturbation Theory in Non-linear Dynamics" in this section.) The coordinate-free approach proves to be quite suitable here. First, let us formalize the observation:

**Lemma 2** *If  $F$  admits a convergent normalizing transformation then there exists a nontrivial  $G$  (i.e.,  $G \notin \mathbb{K} \cdot F$ ) such that  $[G, F] = 0$ .*

**Proof** There exist a convergent  $\psi$  and a convergent  $F^*$  in normal form such that

$$F^* \xrightarrow{\psi} F$$

Now  $\psi$  sends  $B_s$  to some analytic  $G = B_s + \dots$ , and  $[B_s, F^*] = 0$  implies  $[G, F] = 0$ . If  $F^* \neq B_s$ , we are done. If  $F^* = B_s$  then take some linear map  $C \in \mathbb{K} \cdot B_s$  that commutes with  $B_s$ , and define  $G$  by  $C \xrightarrow{\psi} G$ .

In other words, if there is a convergent normalizing transformation then there exists a nontrivial

infinitesimal symmetry at the stationary point. One can try to turn this necessary condition around and thus obtain sufficient convergence criteria. This has been done, with some success, since the early 1990s. Among the relevant contributions are those by Markhashov (1974), Bruno and Walcher (1994), Cicogna (1996, 1997), Bambusi et al. (1998), and, from a different starting point, Stolovitch (2000) (see below and see also Gramchev and Yoshino (1999) for maps). The survey paper (2002) by Cicogna and Walcher collects the development up to 2001.

Here we will present a few results to give the reader an impression of the arguments employed.

The objects to deal with are  $C_{\text{for}}(F)$  and  $C_{\text{an}}(F)$ , i.e., the formal, respectively analytic, centralizer of  $F$ , which by definition consist of all formal, respectively analytic, vector fields  $H$  such that  $[H, F] = 0$ . The first result is due to Cicogna (1996, 1997) and Walcher (2000). Note that Pliss' theorem (and thus Condition A) plays a crucial role in the proof.

**Theorem 6** *Given the analytic vector field  $F$  with formal normal form  $\hat{F}$ , assume that*

$$\dim C_{\text{for}}(\hat{F}) = k < \infty.$$

*If the eigenvalues of  $B$  satisfy Condition  $\omega$  and  $\dim C_{\text{an}}(F) = k$  then there exists a convergent transformation to normal form.*

**Proof** We have  $\dim C_{\text{for}}(F) = \dim C_{\text{for}}(\hat{F})$ , so  $\dim C_{\text{an}}(F) = k$  implies  $C_{\text{an}}(F) = C_{\text{for}}(F)$ . Given a formal transformation  $\psi$  with  $\hat{F} \xrightarrow{\psi} F$ , there exists an analytic  $H$  such that  $B \xrightarrow{\psi} H$  and  $[H, F] = 0$ , since  $B \in C_{\text{for}}(\hat{F})$ . Note that  $H = B + \dots$ , so  $B$  is a normal form of  $H$ . Due to Pliss' theorem, there is a convergent  $\Phi$  with  $B \xrightarrow{\Phi} H$ . Now  $\hat{F} \xrightarrow{\Phi} F$  for some  $\tilde{F}$ , and  $[B, \tilde{F}] = 0$ , so  $\tilde{F}$  is in normal form.  $\square$

The dimension of the formal centralizer is computable in many cases. The requirement on Condition  $\omega$  can be relaxed; see Cicogna (1996), Walcher (2000), and Cicogna and Walcher (2002). The next result is due to Markhashov (1974) for the non-resonant case, and to Bruno

and Walcher (1994) in the resonant case. One may base a proof on the observation that  $\dim C_{\text{for}}(F)$  is infinite only if Condition A holds, and use Theorem 6 otherwise.

**Theorem 7** *In dimension  $n = 2$ , there is a convergent transformation of  $F$  to normal form if and only if  $F$  admits a nontrivial commuting vector field in 0.*

In dimension two, there are other, very precise, characterizations of resonant vector fields (for  $\lambda_2/\lambda_1$  a negative rational number) admitting a convergent normalization, which can be drawn from the work of Martinet and Ramis (1983). Beyond this, building on work of Ecalle (1981a, b) and Voronin (1981), Martinet and Ramis succeeded in giving an analytical classification of germs of such vector fields, and those admitting a convergent normalizing transformation can be characterized by the vanishing of infinitely many analytical invariants. In this sense, the convergence problem was settled earlier, at least for the interesting cases. But Theorem 7 approaches the question from a different perspective, gives an algebraic characterization and provides structural insight that is not directly available from Martinet and Ramis (1983).

Lie algebra arguments also play a fundamental role, from a different perspective, in the work of Stolovitch (2000). We will here give a simplified and incomplete account of his important results; a full presentation would take up much more space. To motivate Stolovitch's approach, note that in many cases there are natural decompositions  $B = B_1 + \dots + B_\ell$ , with all  $[B_i, B_j] = 0$  which come from eigenvalues splitting up into groups of pairwise commensurable ones, with pairwise incommensurability between the groups. The resonance conditions  $\sum m_j \lambda_j - \lambda_k = 0$  then split up into conditions involving only the groups. We illustrate this with a simple example for such a decomposition:

$$\begin{aligned} \text{diag}(1, -1, \sqrt{2}, -\sqrt{2}) &= \text{diag}(1, -1, 0, 0) \\ &+ \text{diag}(0, 0, \sqrt{2}, -\sqrt{2}). \end{aligned}$$

The setting considered by Stolovitch is as follows: Given (complex) analytic vector fields

$$F_1 = B_1 + \dots, \dots, F_\ell = B_\ell + \dots$$

that commute pairwise, thus all  $[F_i, F_j] = 0$ , ask about simultaneous analytic normalization of these vector fields. There are sensible notions of normal form of a vector field with respect to a linear Lie algebra, and in particular there is a natural extension of Poincaré–Dulac for abelian Lie algebras of diagonal matrices; see Stolovitch (2000), section “Convergence and Convergence Problems”: Each semisimple linear part  $B_{i,s}$  commutes with each  $F_j$  in such a normal form. To avoid trivial scenarios, Stolovitch requires the semisimple parts  $B_{i,s}$  to be linearly independent. To formulate diophantine conditions extending Bruno’s Condition  $\omega$ , one may proceed as follows: Let  $\lambda_{i,1}, \dots, \lambda_{i,n}$  denote the eigenvalues of  $B_i$ . For nonnegative integers  $m_1, \dots, m_n$ , and for  $1 \leq d \leq n$  set

$$\gamma_{m_1, \dots, m_n, d} := \sum_i \left| \sum_j m_j \lambda_{i,j} - \lambda_d \right|$$

and (for instance; Stolovitch gives a more general formulation)

$$\omega_k(B_1, \dots, B_\ell) := \inf \left\{ \gamma_{m_1, \dots, m_n, d} \neq 0 : 1 \leq d \leq n, 2 \leq \sum m_j \leq 2^k \right\}.$$

This leads to an appropriate diophantine condition which extends Bruno’s Condition  $\omega$ .

Condition  $\omega^\#$ :

$$-\sum_{k=1}^{\infty} \frac{\omega_k(B_1, \dots, B_\ell)}{2^k} < \infty$$

We remark briefly that there appear to be some issues of well-definedness here. For instance, the choice of the  $F_i$ , and hence of the  $B_i$ , is not unique, and one has to verify that the important notions, like Condition  $\omega^\#$ , do not depend on these choices. Stolovitch (2000) works in an invariant setting from the start; as a consequence, no such questions arise.

Finally, Stolovitch introduces the notion of *formal complete integrability*. Disregarding technical

subtleties (even though these are relevant and of interest), one may informally characterize this property as follows: The system  $F_1, \dots, F_\ell$  is formally completely integrable if it has as many formal integrals as admissible by the semisimple linear parts  $B_{1,s}, \dots, B_{\ell,s}$ . To cast at least some light on this, we note that every formal integral of the system in normal form is also a simultaneous first integral of the  $B_{i,s}$ . See Walcher (1991) for the case  $\ell = 1$  and Stolovitch (2000), section “Lie Algebra Arguments” for the general case. This notion of complete integrability is the appropriate extension of Bruno’s Condition A.

The following is a simplified representative of the results in Stolovitch (2000). The system  $B_1, \dots, B_\ell$  is said to have *small divisors* if 0 is a limit point of the  $\gamma_{m_1, \dots, m_n, d} \neq 0$ .

**Theorem 8 (Stolovitch)** *In the presence of small divisors, if Condition  $\omega^\#$  holds then every formally completely integrable system  $F_1, \dots, F_\ell$  admits a convergent transformation to normal form.*

The principal value of Stolovitch’s results is that they open up a unified approach to a number of applications. One almost immediate application is a recovery of Bruno’s convergence theorem. (See also Remark (c) following Theorem 5.) We give two more applications. For the first one, compare also Bambusi et al. (1998).

**Corollary 1 (Linearization)** *In the presence of small divisors, assuming that Condition  $\omega^\#$  holds, every formally linearizable system  $F_1, \dots, F_\ell$  admits a convergent linearization.*

The second application is a short and clear proof of a theorem due to Vey (1979):

**Corollary 2 (Volume-preserving vector fields)**

*Assume that  $\ell = n - 1$  and that  $F_1, \dots, F_{n-1}$  are commuting volume-preserving vector fields, with diagonal and linearly independent linear parts  $B_i = B_{i,s}$ . Then the  $F_i$  are simultaneously analytically normalizable.*

**Remark (a)** For similar results see Zung (2002). Zung uses a different approach based on a convergence result (a precursor of Zung 2005) for

Hamiltonian vector fields; see also the section on Hamiltonian vector fields below. His argument seems somewhat sketchy.

**Remark (b)** There is a nontrivial overlap of Stolovitch's results with the approach to convergence via symmetries outlined above: Some results can be proven by either method, as the references indicate.

We finish this section with a third aspect of Lie algebra arguments in convergence proofs. The following is taken from Walcher (1993):

**Theorem 9** *Let  $\mathcal{L}$  be a finite dimensional Lie algebra of polynomial vector fields which is graded in the sense that it contains all homogeneous parts of each of its elements, and let  $F \in \mathcal{L}$  and  $F(0) = 0$ . Then  $F$  admits a convergent transformation to normal form  $F^*$ , and one may take  $F^* \in \mathcal{L}$ .*

The basic idea for the proof is to take suitable solutions  $h_r$  of the homological equations, which can be chosen in  $\mathcal{L}$ , and transformations  $\exp(h_r)$ ; see Walcher (1993, Prop. 2.7). Due to finite dimension of  $\mathcal{L}$ , finitely many such transformations suffice. There are several interesting Lie algebras among the finite dimensional graded ones:

**Example (a)** Every projective vector field, as well as every conformal vector field in dimension  $\geq 3$ , admits a convergent transformation to Poincaré–Dulac normal form.

**Example (b)** Matrix Riccati equations: Every matrix differential equation of the form

$$\dot{x} = xax + bx + xc,$$

with  $x, a, b, c$  matrices of appropriate sizes, admits a convergent transformation to normal form.

## NFIM and Sets of Analyticity

As we have seen, convergence problems for normalizing transformations are unsurmountable in many instances. But there are sensible strategies

to achieve convergence by relaxing the requirements on the normalized vector field. Moreover, such strategies frequently provide interesting information, e.g., on stability or on the existence of periodic solutions. There are two related, but not equivalent, approaches to be discussed here: Bibikov's *normal form on an invariant manifold (NFIM)*, see Bibikov (1979); and Bruno's *sets of analyticity*, see Bruno (1989).

We will first discuss Bibikov's work, including a coordinate-free approach proposed in Walcher (1993).

**Definition 3** Let  $C$  be a semisimple linear map. A vector subspace of  $\mathbb{K}^n$  is called *strongly  $C$ -stable* if it is invariant for every vector field  $F = B + \dots$ , with  $B_s = C$ , in normal form.

Strongly  $C$ -stable spaces are in particular  $C$ -stable. A coordinate-dependent characterization is as follows.

**Lemma 3** *Assume that  $C$  is in diagonal form, and let  $1 < r < n$ . Then the subspace  $U := \mathbb{K}e_1 + \dots + \mathbb{K}e_r$  is strongly  $C$ -stable if and only if*

$$m_1\lambda_1 + \dots + m_r\lambda_r - \lambda_j \neq 0$$

*for all nonnegative integers  $m_i$  with  $\sum m_i > 0$ , and all  $j \in \{r+1, \dots, n\}$ .*

Here, the choice of indices  $1, \dots, r$  is just for the sake of convenience. The proof is simple: The condition ensures that no monomial

$$x_1^{m_1} \dots x_r^{m_r} e_j, \quad j > r$$

will occur in the normal form, and therefore the subspace  $U$ , characterized by  $x_{r+1} = \dots = x_n = 0$ , is invariant for any normal form  $F = C + B_n + \dots$ . Examples include the stable, unstable and center subspaces of  $C$ .

Now one can introduce the notion of NFIM:

**Definition 4** Assume that  $U = \mathbb{K}e_1 + \dots + \mathbb{K}e_r$  is strongly  $C$ -stable. A vector field  $F = B + \dots$  with  $B_s = C$  is said to be in *normal form on the invariant manifold  $U$  (NFIM on  $U$ )* if  $U$  is invariant for  $F$  and furthermore

$$[B_s|_U, F|_U] = 0.$$

**Example** The two-dimensional vector field

$$F = \begin{pmatrix} x_1^2 - 2x_1x_2 + 3x_1^3 \\ x_2(1 + x_1 + x_2^2) \end{pmatrix}, \text{ with } Bx = \begin{pmatrix} 0 \\ x_2 \end{pmatrix},$$

is in NFIM on  $U = \mathbb{K}e_1$ .

Bibikov also introduced a refined version of NFIM, which he calls quasi-normal form (QNF). The above example is not in quasi-normalform; in a QNF the first entry would contain only functions of  $x_1$ .

Because normal forms are, in particular, in NFIM on every strongly  $B_s$ -stable subspace, it is obvious that formal transformations to NFIM exist. But there is more freedom to construct such transformations, which may be utilized to force convergence. To give a flavor of Bibikov's results, we present a weakened version of one of his theorems.

**Theorem 10 (Bibikov)** *Given a vector field  $F = B + \dots$ , assume that the subspace  $U := \mathbb{K}e_1 + \dots + \mathbb{K}e_r$  is strongly  $B_s$ -stable, and moreover that:*

- (i) *There is an  $\epsilon > 0$  such that*

$$|m_1\lambda_1 + \dots + m_r\lambda_r - \lambda_j| \geq \epsilon$$

*for all nonnegative integers  $m_i$ ,  $\sum m_i > 0$ , and for all  $j > r$ .*

- (ii) *Some formal normal form  $\hat{F}|_U$  satisfies the Pliss condition on  $U$ . Then there exists a convergent transformation to NFIM on  $U$ .*

**Remark** In Bibikov's original theorems (see Bibikov 1979, Theorems 3.2 and 10.2), one finds a more general (quite technical) condition instead of (ii). Thus the range of Bibikov's theorems is wider than our statement indicates. Condition (i), or some related condition, cannot be discarded completely: There are examples of strongly  $C$ -stable subspaces which do not correspond to analytic invariant manifolds for certain

vector fields. (This is another incarnation of the small denominator problem.)

Applications of Bibikov's theorems include the existence of analytic stable and unstable manifolds; stability of the stationary point in case  $\lambda_1 = 0$  when all other  $\lambda_i$  have negative real parts, and  $\hat{F}|_U = 0$  on  $U = \mathbb{K}e_1$  ("transcendental case"); and the existence of certain periodic solutions.

Let us now turn to Bruno's method of analytic invariant sets; see Bruno Part I, Chap. III, and Part II in Bruno (1989). To motivate the approach, one may use the following observation: For an analytic vector field  $\hat{F} = B + \dots$  in normal form the set

$$\mathcal{A} = \left\{ z : \hat{F}(z) \text{ and } B_s z \text{ linearly dependent in } \mathbb{K}^n \right\}$$

is invariant for  $\hat{F}$ . (This is a consequence of  $[B_s, \hat{F}] = 0$ : The set of points for which a vector field and an infinitesimal symmetry "point in the same direction" is invariant.) It is clearly possible to write down analytic functions such that  $\mathcal{A}$  is their common zero set: For instance, take suitable  $2 \times 2$ -determinants. One may now introduce the notion that a vector field  $\hat{F}$  is *normalized on  $\mathcal{A}$* : In the coordinate version this means that for each entry of the right-hand side the sum of all nonresonant parts lies in the defining ideal of  $\mathcal{A}$ .

If  $F$  is not in normal form but some formal normal form  $\hat{F}$  satisfies Condition A then – assuming some mild diophantine conditions –  $\mathcal{A}$  is a whole neighborhood of the stationary point, according to Bruno's convergence theorem. Bruno now refines this by investigating whether the generally "formal" set  $\mathcal{A}$  is analytic (or at least certain subsets are), and what can be said about the solutions on such subsets. To be more precise: Given a not necessarily convergent normal form  $\hat{F}$  of  $F$ , one can still write down formal power series whose "common zero set" defines  $\mathcal{A}$ , and it makes sense to ask what of this can be salvaged for analyticity. The following is a sample of Bruno's results (see Part I, Chap. III, Theorems 2 and 4 in Bruno (1989)). As above, we do not write down the technical conditions completely.



**Theorem 11 (Bruno)** *Let the analytic vector field  $F = B + \dots$  be given.*

- (a) *If the eigenvalues  $\lambda_1, \dots, \lambda_n$  of  $B$  are commensurable (i.e., pairwise linearly dependent over the rationals) then the set  $\mathcal{A}$  is analytic and there is a convergent transformation to a vector field  $\tilde{F}$  that is normalized on  $\mathcal{A}$ .*
- (b) *Generally, there exists a (formal) subset  $\mathcal{B}$  of  $\mathcal{A}$  which is analytic, and for  $\mathcal{B}$  the same conclusion as above holds.*

For applications see Bruno (1989), Part II, and the recent papers by Edneral (2005), and Bruno and Edneral (2006), on existence of periodic solutions for certain equations.

## Hamiltonian Systems

Hamiltonian systems have a special position among differential equations (see e.g., Arnold et al. (1993) for an overview, and ▶ “Hamiltonian Perturbation Theory (and Transition to Chaos)”; in view of their importance it is appropriate to give them particular attention. When discussing normal forms of Hamiltonian systems (where Poincaré–Dulac becomes Birkhoff; see 1927) it is natural to consider canonical transformations only. In view of the correspondence between integrals of  $F$  and Hamiltonian vector fields commuting with  $F$ , it is furthermore natural to consider integrals in the Hamiltonian setting. As for convergence results, we first state two theorems by H. Ito (1989, 1992), from around 1990:

**Theorem 12 (Ito; non-resonant case)** *Let  $F = B + \dots$  be Hamiltonian and let  $(\omega_1, -\omega_1, \dots, \omega_r, -\omega_r)$  be the eigenvalues of  $B$ . Moreover assume that  $\omega_1, \dots, \omega_r$  are non-resonant, thus  $\sum m_j \omega_j = 0$  for integers  $m_1, \dots, m_r$  implies  $m_1 = \dots = m_r = 0$ . If  $F$  possesses  $r$  independent integrals in involution (i.e. with vanishing Poisson brackets) then there exists a convergent canonical transformation of  $f$  to Birkhoff normal form.*

This condition is also necessary in the non-resonant case: If there is a convergent

transformation to analytic normal form  $\hat{F}$  then there are  $r$  linearly independent linear Hamiltonian vector fields that commute with  $\hat{F}$ , and these, in turn, correspond to  $r$  independent quadratic integrals of  $\hat{F}$ . For the “single resonance” case one has:

**Theorem 13 (Ito; a simple-resonance case)** *Let  $F = B + \dots$  be Hamiltonian and let  $(\omega_1, -\omega_1, \dots, \omega_r, -\omega_r)$  be the eigenvalues of  $B$ . Moreover assume that there are nonzero integers  $n_1, n_2$  such that  $n_1 \omega_1 + n_2 \omega_2 = 0$ , but there are no further resonances. If  $F$  possesses  $r$  independent integrals in involution then there exists a convergent canonical transformation to Birkhoff normal form.*

Again, the condition is also necessary. Ito’s proofs are quite long and intricate. Kappeler, Kodama, and Nemethi (1998) proved a generalization of Theorem 13 for more general single-resonance cases. Moreover, they showed that there is a natural obstacle to further generalizations, since there exist non-integrable (polynomial) Hamiltonian systems in normal form. Thus, a complete integrability condition is not generally necessary for convergence.

Nguyen Tien Zung (2005) recently succeeded in a far-reaching generalization of Ito’s theorems. Considering the existence of non-integrable normal forms, this seems the best possible result on integrability and convergence.

**Theorem 14 (Zung)** *Any analytically integrable Hamiltonian system near a stationary point admits a convergent transformation to Birkhoff normal form.*

One remarkable feature of Zung’s proof is its relative shortness, compared with the proofs by Ito.

Finally, turning to divergent normal forms (rather than normalizing transformations) of Hamiltonian systems, Perez-Marco (2003) recently established a theorem about convergence or generic divergence of the normal form in the non-resonant scenario. Although numerical computations indicate the existence of analytic Hamiltonian vector fields which admit only divergent normal forms, there still seems to be no example known. Perez–Marco showed that if there is one example then divergence is generic.



## Future Directions

There are various ways to extend the approaches and results presented above, and clearly there are unresolved questions. In the following, some of these will be listed, respectively, recalled.

The convergence problem for normal forms and normalizing transformations is part of the much bigger problem of analytic classification of germs of vector fields. Except for dimension two (see the references Ecalle (1981a, b) and Martinet and Ramis (1983)) mentioned above) little seems to be known in the case of nontrivial formal normal forms. Going beyond analyticity, an interesting extension would be towards Gevrey spaces. Some work on this topic exists already; see e.g., Gramchev and Tolis (2006).

Passing from vector fields to maps, matters turn out to be much more complicated in the case of nontrivial formal normal forms, and even one-dimensional maps show very rich behavior; see Perez-Marco (1995, 1997). A brief introduction is given in the survey (Gramchev and Walcher 2005) mentioned above.

A complete (“algebraic”) understanding of Bruno’s Condition A would be desirable; this could also provide an approach to a non-Hamiltonian version of Zung’s Theorem 14. It seems well possible that all the necessary ingredients for this endeavor are contained in Stolovitch’s work (2000).

An extension or refinement of Bibikov’s and Bruno’s results on the existence of certain invariant sets (in the case of non-convergence) would obviously be interesting. There seems to be a natural guideline here: Check what invariant sets are forced onto a system in PDNF and see which ones can be salvaged. (The existence of a commuting vector field, for instance, has more consequences than those exploited by Bruno in the arguments leading up to Theorem 11.)

Finally, one could turn to more refined versions of normal forms, such as normal forms with respect to a nilpotent linear part (see Cushman and Sanders 1990), and quite general constructions such as presented by Sanders (2003, 2005). It seems that little attention has been paid to convergence questions for such types of normal

forms. For normal forms with respect to a nilpotent linear part, there are obviously no small denominator problems, but algebraic obstructions abound. There exists a precise algebraic characterization for such normal forms, involving the representation theory of  $sl(2)$  (see Cushman and Sanders 1990). Thus there may be some hope for an algebraic characterization of convergently normalizable vector fields.

## Bibliography

- Arnold VI (1982) Geometrical methods in the theory of ordinary differential equations. Springer, Berlin
- Arnold VI, Kozlov VV, Neishtadt AI (1993) Mathematical aspects of classical and celestial mechanics. In: Arnold VI (ed) Dynamical systems III, Encyclopaedia mathematical sciences, vol 3, 2nd edn. Springer, New York
- Bambusi D, Cicogna G, Gaeta G, Marmo G (1998) Normal forms, symmetry and linearization of dynamical systems. *J Phys A* 31:5065–5082
- Bibikov YN (1979) Local theory of nonlinear analytic ordinary differential equations, Lecture notes in math 702. Springer, New York
- Birkhoff GD (1927) Dynamical systems, vol IX. American Mathematical Society, Colloquium Publications, Providence
- Bruno AD (1971) Analytical form of differential equations. *Trans Mosc Math Soc* 25:131–288
- Bruno AD (1989) Local methods in nonlinear differential equations. Springer, New York
- Bruno AD, Edneral VF (2006) The normal form and the integrability of systems of ordinary differential equations (Russian). *Programmirovaniye* 3:22–29. (Translation in *Program Comput Softw* 32(3):139–144)
- Bruno AD, Walcher S (1994) Symmetries and convergence of normalizing transformations. *J Math Anal Appl* 183: 571–576
- Chow S-N, Li C, Wang D (1994) Normal forms and bifurcations of planar vector fields. Cambridge University Press, Cambridge, UK
- Cicogna G (1996) On the convergence of normalizing transformations in the presence of symmetries. *J Math Anal Appl* 199:243–255
- Cicogna G (1997) Convergent normal forms of symmetric dynamical systems. *J Phys A* 30:6021–6028
- Cicogna G, Gaeta G (1999) Symmetry and perturbation theory in nonlinear dynamics, Lecture notes in physics monographs 57. Springer, New York
- Cicogna G, Walcher S (2002) Convergence of normal form transformations: the role of symmetries. *Acta Appl Math* 70:95–111
- Cushman R, Sanders JA (1990) A survey of invariant theory applied to normal forms of vector fields with nilpotent linear part, *IMA vol math appl* 19. Springer, New York, pp 82–106

- DeLatté D, Gramchev T (2002) Biholomorphic maps with linear parts having Jordan blocks: linearization and resonance type phenomena. *Math Phys Electron J* 8(2):27. (Electronic)
- Dulac H (1912) Solutions d'un système d'équations différentielles dans le voisinage de valeurs singulières. *Bull Soc Math Fr* 40:324–383
- Ecalte J (1981a) Sur les fonctions résurgentes I. *Publ Math d'Orsay* 81(5):1–247
- Ecalte J (1981b) Sur les fonctions résurgentes II. *Publ Math d'Orsay* 81(6):248–531
- Edneral VF (2005) Looking for periodic solutions of ODE systems by the normal form method. In: *Differential equations with symbolic computation*, Trends math. Birkhäuser, Basel, pp 173–200
- Gramchev T (2002) On the linearization of holomorphic vector fields in the Siegel domain with linear parts having nontrivial Jordan blocks SPT. In: *Symmetry and perturbation theory* (Cala Gonone). World Scientific Publishing, River Edge, pp 106–115
- Gramchev T, Tolis E (2006) Solvability of systems of singular partial differential equations in function spaces. *Integral Transforms Spec Funct* 17: 231–237
- Gramchev T, Walcher S (2005) Normal forms of maps: formal and algebraic aspects. *Acta Appl Math* 87(1–3, 123):–146
- Gramchev T, Yoshino M (1999) Rapidly convergent iteration method for simultaneous normal forms of commuting maps. *Math Z* 231:745–770
- Iooss G, Adelmeyer M (1992) *Topics in bifurcation theory and applications*. World Scientific, Singapore
- Ito H (1989) Convergence of Birkhoff normal forms for integrable systems. *Comment Math Helv* 64:412–461
- Ito H (1992) Integrability of Hamiltonian systems and Birkhoff normal forms in the simple resonance case. *Math Ann* 292:411–444
- Kappeler T, Kodama Y, Nemethi A (1998) On the Birkhoff normal form of a completely integrable system near a fixed point in resonance. *Ann Scuola Norm Sup Pisa Cl Sci* 26:623–661
- Markhashov LM (1974) On the reduction of an analytic system of differential equations to the normal form by an analytic transformation. *J Appl Math Mech* 38: 788–790
- Martinet J, Ramis J-P (1983) Classification analytique des équations différentielles non linéaires résonnantes du premier ordre. *Ann Sci Ecole Norm Sup* 16:571–621
- Perez Marco R (1995) Nonlinearizable holomorphic dynamics having an uncountable number of symmetries. *Invent Math* 119:67–127
- Perez-Marco R (1997) Fixed points and circle maps. *Acta Math* 179:243–294
- Perez-Marco R (2003) Convergence and generic divergence of the Birkhoff normal form. *Ann Math* 157: 557–574
- Pliss VA (1965) On the reduction of an analytic system of differential equations to linear form. *Differ Equ* 1: 153–161
- Poincaré H (1879) *Sur les propriétés des fonctions définies par les équations aux différences partielles*. These, Paris
- Sanders JA (2003) Normal form theory and spectral sequences. *J Differ Equ* 192:536–552
- Sanders JA (2005) Normal form in filtered Lie algebra representations. *Acta Appl Math* 87:165–189
- Siegel CL (1952) Über die Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung, vol 23. *Nachr Akad Wiss Göttingen Math Phys Kl*, pp 21–30
- Stolovitch L (2000) Singular complete integrability. *IHES Publ Math* 91:133–210
- Verhulst F (2005) *Methods and applications of singular perturbations: boundary layers and multiple timescale dynamics*, Texts in applied mathematics, vol 50. Springer, New York
- Vey J (1979) Algèbres commutatives de champs de vecteurs isochores. *Bull Soc Math Fr* 107(4):423–432
- Voronin SM (1981) Analytic classification of germs of conformal mappings  $(C,0) \rightarrow (C,0)$ . *Funct Anal Appl* 15:1–13
- Walcher S (1991) On differential equations in normal form. *Math Ann* 291:293–314
- Walcher S (1993) On transformations into normal form. *J Math Anal Appl* 180(2):617–632
- Walcher S (2000) On convergent normal form transformations in presence of symmetries. *J Math Anal Appl* 244: 17–26
- Zung NT (2002) Convergence versus integrability in Poincaré–Dulac normal form. *Math Res Lett* 9(2–3):217–228
- Zung NT (2005) Convergence versus integrability in Birkhoff normal form. *Ann Math* 161(1):141–156

# Diagrammatic Methods in Classical Perturbation Theory

Guido Gentile

Dipartimento di Matematica, Università di Roma Tre, Rome, Italy

## Article Outline

Glossary

Definition of the Subject

Introduction

Examples

Trees and Graphical Representation

Small Divisors

Multiscale Analysis

Generalizations

Conclusions and Future Directions

Bibliography

## Glossary

**Dynamical system** Let  $W \subseteq \mathbb{R}^N$  be an open set and  $f: W \times \mathbb{R} \rightarrow \mathbb{R}^N$  be a smooth function. The ordinary differential equation  $\dot{x} = f(x, t)$  on  $W$  defines a continuous dynamical system. A discrete dynamical system on  $W$  is defined by a map  $x \rightarrow x' = F(x)$ , with  $F$  depending smoothly on  $x$ .

**Hamiltonian system** Let  $\mathcal{A} \subseteq \mathbb{R}^d$  be an open set and  $\mathcal{H}: \mathcal{A} \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$  be a smooth function ( $\mathcal{A} \times \mathbb{R}^d$  is called the phase space). Consider the system of ordinary differential equations  $\dot{q}_k = \partial \mathcal{H}(q, p, t) / \partial p_k$ ,  $\dot{p}_k = -\partial \mathcal{H}(q, p, t) / \partial q_k$ , for  $k = 1, \dots, d$ . The equations are called Hamilton equations, and  $\mathcal{H}$  is called a Hamiltonian function. A dynamical system described by Hamilton equations is called a Hamiltonian system.

**Integrable system** A Hamiltonian system is called integrable if there exists a system of coordinates  $(\alpha, A) \in \mathbb{T}^d \times \mathbb{R}^d$ , called angle-action variables, such that in these coordinates the motion is  $(\alpha, A) \rightarrow (\alpha + \omega(A)t, A)$ , for some

smooth function  $\omega(A)$ . Hence in these coordinates the Hamiltonian function  $\mathcal{H}$  depends only on the action variables,  $\mathcal{H} = \mathcal{H}_0(A)$ .

**Invariant torus** Given a continuous dynamical system we say that the motion occurs on an invariant  $d$ -torus if it takes place on a  $d$ -dimensional manifold and its position on the manifold is identified through a coordinate in  $\mathbb{T}^d$ . In an integrable Hamiltonian system all phase space is filled by invariant tori. In a quasi-integrable system the KAM theorem states that most of the invariant tori persist under perturbation, in the sense that the relative Lebesgue measure of the fraction of phase space filled by invariant tori tends to 1 as the perturbation tends to disappear. The persisting invariant tori are slight deformations of the unperturbed invariant tori.

**Quasi-integrable system** A quasi-integrable system is a Hamiltonian system described by a Hamiltonian function of the form  $\mathcal{H} = \mathcal{H}_0(A) + \varepsilon f(\alpha, A)$ , with  $(\alpha, A)$  angle-action variables,  $\varepsilon$  a small real parameter and  $f$  periodic in its arguments  $\alpha$ .

**Quasi-periodic motion** Consider the motion  $\alpha \rightarrow \alpha + \omega t$  on  $\mathbb{T}^2$ , with  $\omega = (\omega_1, \omega_2)$ . If  $\omega_1/\omega_2$  is rational, the motion is periodic, that is there exists  $T > 0$  such that  $\omega_1 T = \omega_2 T = 0 \bmod 2\pi$ . If  $\omega_1/\omega_2$  is irrational, the motion never returns to its initial value. On the other hand it densely fills  $\mathbb{T}^2$ , in the sense that it comes arbitrarily close to any point of  $\mathbb{T}^2$ . We say in that case that the motion is quasi-periodic. The definition extends to  $\mathbb{T}^d$ ,  $d > 2$ : a linear motion  $\alpha \rightarrow \alpha + \omega t$  on  $\mathbb{T}^d$  is quasi-periodic if the components of  $\omega$  are rationally independent, that is if  $v = \omega_1 v_1 + \dots + \omega_d v_d = 0$  for  $v \in \mathbb{Z}^d$  if and only if  $v = 0$  ( $a \cdot b$  is the standard scalar product between the two vectors  $a, b$ ). More generally we say that a motion on a manifold is quasi-periodic if, in suitable coordinates, it can be described as a linear quasi-periodic motion. The vector  $\omega$  is usually called the frequency or rotation vector.

**Renormalization group** By renormalization group one denotes the set of techniques and concepts used to study problems where there are some scale invariance properties. The basic mechanism consists in considering equations depending on some parameters and defining some transformations on the equations, including a suitable rescaling, such that after the transformation the equations can be expressed, up to irrelevant corrections, in the same form as before but with new values for the parameters.

**Torus** The 1-torus  $\mathbb{T}$  is defined as  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ , that is the set of real numbers defined modulo  $2\pi$  (this means that  $x$  is identified with  $y$  if  $x - y$  is a multiple of  $2\pi$ ). So it is the natural domain of an angle. One defines the  $d$ -torus  $\mathbb{T}^d$  as a product of  $d$  1-tori, that is  $\mathbb{T}^d = \mathbb{T} \times \dots \times \mathbb{T}$ . For instance one can imagine  $\mathbb{T}^2$  as a square with the opposite sides glued together.

**Tree** A graph is a collection of points, called nodes, and of lines which connect the nodes. A walk on the graph is a sequence of lines such that any two successive lines in the sequence share a node; a walk is nontrivial if it contains at least one line. A tree is a planar graph with no closed loops, that is, such that there is no non-trivial walk connecting any node to itself. An oriented tree is a tree with a special node such that all lines of the tree are oriented toward that node. If we add a further oriented line connecting the special node to another point, called the root, we obtain a rooted tree (see Fig. 1 in section “Trees and Graphical Representation”).

## Definition of the Subject

Recursive equations naturally arise whenever a dynamical system is considered in the regime of perturbation theory; for an introductory article on perturbation theory see ▶ “Perturbation Theory”. A classical example is provided by Celestial Mechanics, where perturbation series, known as Lindstedt series, are widely used; see Gallavotti (2006) and ▶ “Perturbation Theory in Celestial Mechanics”.

A typical problem in Celestial Mechanics is to study formal solutions of given ordinary differential equations in the form of expansions in a suitable

small parameter, the perturbation parameter. In the case of quasi-periodic solutions, the study of the series, in particular of its convergence, is made difficult by the presence of the small divisors – which will be defined later on. Under some non-resonance condition on the frequency vector, one can show that the series are well-defined to any order. The first proof of such a property was given by Poincaré (1892–1899), even if the convergence of the series remained an open problem up to the advent of KAM theory – an account can be found in Gallavotti (1983) and in Arnold et al. (1988); see also ▶ “Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite Dimensional Systems”. KAM is an acronym standing for Kolmogorov (1954), Arnold (1963), and Moser (1962), who proved in the middle of last century the persistence of most of invariant tori for quasi-integrable systems.

Kolmogorov and Arnold proofs apply to analytic Hamiltonian systems, while Moser’s approach deals also with the differentiable case; the smoothness condition on the Hamiltonian function was thereafter improved by Pöschel (1982). In the analytic case, the persisting tori turn out to be analytic in the perturbation parameter, as explicitly showed by Moser (1967). In particular, this means that the perturbation series not only are well-defined, but also converge. However, a systematic analysis with diagrammatic techniques started only recently after the pioneering, fundamental works by Eliasson (1996) and Gallavotti (1994a), and were subsequently extended to many other problems with small divisors, including dynamical systems with infinitely many degrees of freedom, such as nonlinear partial differential equations, and non-Hamiltonian systems. Some of these extensions will be discussed in section “Generalizations”.

From a technical point of view, the diagrammatic techniques used in classical perturbation theory are strongly reminiscent of the Feynman diagrams used in quantum field theory: this was first pointed out by Gallavotti (1994a). Also the multiscale analysis used to control the small divisors is typical of renormalization group techniques, which have been successfully used in problems of quantum field theory, statistical mechanics and classical mechanics; see Gallavotti (2001) and Gentile & Mastropietro (2001) for some reviews.

In this section we consider a few paradigmatic examples of dynamical systems which can be described by equations of the form (1).

### A Class of Quasi-integrable Hamiltonian Systems

Consider the Hamiltonian system described by the Hamiltonian function

$$\mathcal{H}(\alpha, A) = \frac{1}{2}A^2 + \varepsilon f(\alpha), \quad (5)$$

where  $(\alpha, A) \in \mathbb{T}^d \times \mathbb{R}^d$  are angle-action variables, with  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ ,  $f$  is a real analytic function,  $2\pi$ -periodic in each of its arguments, and  $A^2 = A \cdot A$ , if (here and henceforth)  $\cdot$  denotes the standard scalar product in  $\mathbb{R}^d$ , that is  $a \cdot b = a_1b_1 + \dots + a_db_d$ . Assume also for simplicity that  $f$  is a trigonometric polynomial of degree  $N$ .

The corresponding Hamilton equations are (we shorten  $\partial_x = \partial/\partial_x$ )

$$\begin{aligned} \dot{\alpha} &= \partial_A \mathcal{H}(\alpha, A) = A \\ \dot{A} &= -\partial_\alpha \mathcal{H}(\alpha, A) = -\varepsilon \partial_\alpha f(\alpha) \end{aligned}$$

which can be written as an equation involving only the angle variables:

$$\ddot{\alpha} = -\varepsilon \partial_\alpha f(\alpha), \quad (6)$$

which is of the form (1) with  $u = \alpha$ ,  $G = 0$ ,  $F = -\partial_\alpha f$ , and  $D = d^2/dt^2$ .

For  $\varepsilon = 0$ ,  $\alpha^{(0)}(t) = \alpha_0 + \omega t$  is a solution of Eq. (6) for any choice of  $\alpha_0 \in \mathbb{T}^d$  and  $\omega \in \mathbb{R}^d$ . Take for simplicity  $\alpha_0 = 0$ : we shall see that this choice makes sense. We say that for  $\varepsilon = 0$  the Hamiltonian function (5) describes a system of *drotators*.

We call  $\omega$  the *frequency vector*, and we say that  $\omega$  is irrational if its components are rationally independent, that is if  $\omega \cdot v = 0$  for  $v \in \mathbb{Z}^d$  if and only if  $v = 0$ . For irrational  $\omega$  the solution  $\alpha^{(0)}(t)$  describes a *quasi-periodic motion* with frequency vector  $\omega$ , and it densely fills  $\mathbb{T}^d$ .

Then Eq. (3) becomes

$$\begin{aligned} \ddot{\alpha}^{(k)} &= -[\varepsilon f(\alpha)]^{(k)} \\ &:= -\sum_{s=0}^{k-1} \frac{1}{s!} \partial_x^{s+1} f(\omega t) \sum_{\substack{k_1 + \dots + k_s = k-1 \\ k_i \geq 1}} \alpha^{(k_1)} \dots \alpha^{(k_s)}. \end{aligned} \quad (7)$$

We look for a quasi-periodic solution of Eq. (6), that is a solution of the form  $\alpha(t) = \omega t +$

$h(\omega t)$ , with  $h(\omega t) = O(\varepsilon)$ . We call  $h$  the *conjugate function*, as it “conjugates” (that is, maps) the perturbed solution  $\alpha(t)$  to the unperturbed solution  $\omega t$ . In terms of the function  $h$  (6) becomes

$$\ddot{h} = -\varepsilon \partial_\alpha f(\omega t + h), \quad (8)$$

where  $\partial_\alpha$  denotes derivative with respect to the argument. Then Eq. (8) can be more conveniently written in Fourier space, where the operator  $D$  acts as a multiplication operator.

If we write

$$h(\omega t) = \sum_{v \in \mathbb{Z}^d} e^{i\omega \cdot vt} h_v, \quad h_v = \sum_{k=1}^{\infty} \varepsilon^k h_v^{(k)}, \quad (9)$$

and insert Eq. (9) into Eq. (8) we obtain

$$\begin{aligned} (\omega \cdot v)^2 h_v^{(k)} &= [\varepsilon \partial_\alpha f(\alpha)]_v^{(k)} \\ &:= \sum_{s=0}^{k-1} \sum_{\substack{k_1 + \dots + k_s = k-1 \\ v_0 + v_1 + \dots + v_s = v \\ v_i \in \mathbb{Z}^d}} \sum_{\substack{v_i \in \mathbb{Z}^d \\ v_0 + v_1 + \dots + v_s = v}} \frac{1}{s!} \\ &\quad \cdot (i v_0)^{s+1} f_{v_0} h_{v_1}^{(k_1)} \dots h_{v_s}^{(k_s)}. \end{aligned} \quad (10)$$

These equations are well-defined to all orders provided  $[\varepsilon \partial_\alpha f(\alpha)]_v^{(k)} = 0$  for all  $v$  such that  $\omega \cdot v = 0$ . If  $\omega$  is an irrational vector we need  $[\varepsilon \partial_\alpha f(\alpha)]_0^{(k)} = 0$  for the equations to be well-defined. In that case the coefficients  $h_0^{(k)}$  are left undetermined, and we can fix them arbitrarily to vanish (which is a convenient choice).

We shall see that under some condition on  $\omega$  a quasi-periodic solution  $\alpha(t)$  exists, and densely fills a  $d$ -dimensional manifold. The analysis carried out above for  $\alpha_0 = 0$  can be repeated unchanged for all values of  $\alpha_0 \in \mathbb{T}^d$ :  $\alpha_0$  represents the *initial phase* of the solution, and by varying  $\alpha_0$  we cover all the manifold. Such a manifold can be parametrized in terms of  $\alpha_0$ , so it represents an invariant torus for the perturbed system.

### A Simplified Model with No Small Divisors

Consider the same equation as (8) with  $D = d^2/dt^2$  replaced by  $-1$ , that is



$$h = \varepsilon \partial_{\alpha} f(\omega t + h). \quad (11)$$

Of course in this case we no longer have a differential equation; still, we can look again for quasi-periodic solutions  $h(\omega t) = O(\varepsilon)$  with frequency vector  $\omega$ . In such a case in Fourier space we have

$$\begin{aligned} h_v^{(k)} &= [\varepsilon \partial_{\alpha} f(\alpha)]_v^{(k)} \\ &:= \sum_{s=0}^{k-1} k_1 \sum_{k_1+\dots+k_s=k-1} \sum_{v_0+v_1+\dots+v_s=v} \frac{1}{s!} (iv_0)^{s+1} \\ &\quad f_{v_0} h_{v_1}^{(k_1)} \dots h_{v_s}^{(k_s)} \end{aligned} \quad (12)$$

For instance if  $d = 1$  and  $f(\alpha) = \cos \alpha$  the equation, which is known as the *Kepler equation*, can be explicitly solved by the Lagrange inversion theorem (Wintner 1941), and gives

$$h_v^{(k)} = \begin{cases} \frac{-i(-1)^{k+(v-k)/2}}{2^k((k-v)/2)!((k+v)/2)!}, & |v| \leq k, \\ 0, & v+k \text{ even,} \\ & \text{otherwise.} \end{cases} \quad (13)$$

We shall show in section “[Small Divisors](#)” that a different derivation can be provided by using the forthcoming diagrammatic techniques.

### The Standard Map

Consider the finite difference equation

$$D\alpha = -\varepsilon \sin \alpha, \quad (14)$$

on  $\mathbb{T}$ , where now  $D$  is defined by

$$D\alpha(\psi) := 2\alpha(\psi) - \alpha(\psi + \omega) - \alpha(\psi - \omega). \quad (15)$$

By writing  $\alpha = \psi + h(\psi)$ , Eq. (14) becomes

$$Dh = -\varepsilon \sin(\psi + h), \quad (16)$$

which is the functional equation that must be solved by the *conjugation function* of the *standard map*

$$\begin{cases} x' = x + y + \varepsilon \sin x, \\ y' = y + \varepsilon \sin x. \end{cases} \quad (17)$$

In other words, by writing  $x = \psi + h(\psi)$  and  $y = \omega + h(\psi) - h(\psi - \omega)$ , with  $(\psi, \omega)$  solving Eq. (17) for  $\varepsilon = 0$ , that is  $(\psi', \omega') = (\psi + \omega, \omega)$ , we obtain a closed-form equation for  $h$ , which is exactly Eq. (16).

In Fourier space the operator  $D$  acts as  $D : e^{iv\psi} \rightarrow 4\sin^2(\omega v/2)e^{iv\psi}$ , so that, by expanding  $h$  according to Eq. (9), we can write Eq. (16) as

$$\begin{aligned} h_v^{(k)} &= \frac{1}{4\sin^2(\omega v/2)} \sum_{s=0}^{k-1} \sum_{k_1+\dots+k_s=k-1} \sum_{v_0+v_1+\dots+v_s=v} \frac{1}{s!} \\ &\quad \cdot (iv_0)^{s+1} f_{v_0} h_{v_1}^{(k_1)} \dots h_{v_s}^{(k_s)}, \end{aligned} \quad (18)$$

where  $v_0 = \pm 1$  and  $f_{\pm 1} = 1/2$ .

Note that Eq. (17) is a discrete dynamical system. However, when passing to Fourier space, Eq. (18) acquires the same form as for the continuous dynamical systems previously considered, simply with a different kernel for  $D$ . In particular if we replace  $D$  with 1 we recover the Kepler equation.

The number  $\omega$  is called the *rotation number*. We say that  $\omega$  is irrational if the vector  $(2\pi, \omega)$  is irrational according to the previous definition.

### Trees and Graphical Representation

Take  $\omega$  to be irrational. We study the recursive equations

$$\begin{cases} h_v^{(k)} = g(\omega \cdot v) [\varepsilon \partial_{\alpha} f(\alpha)]_v^{(k)}, & v \neq 0, \\ [\varepsilon \partial_{\alpha} f(\alpha)]_0^{(k)} = 0, & v = 0, \end{cases} \quad (19)$$

where the form of  $g$  depends on the particular model we are investigating. Hence one has either  $g(\omega \cdot v) = (\omega \cdot v)^{-2}$  or  $g(\omega \cdot v) = 1$  or  $g(\omega \cdot v) = (2\sin(\omega v/2))^{-2}$  according to models described in section “[Examples](#)”.

For  $v \neq 0$  we have equations which express the coefficients  $h_v^{(k)}$ ,  $v \in \mathbb{Z}^d$ , in terms of the coefficients



$h_v^{(k)}$ ,  $v \in \mathbb{Z}^d$ , with  $k' < k$ , provided the equations for  $v = 0$  are satisfied for all  $k \geq 1$ . Recursive equations, such as Eq. (19), naturally lead to a graphical representation in terms of trees.

### Trees

A connected *graph*  $\mathcal{G}$  is a collection of points (nodes) and lines connecting all of them. Denote with  $N(\mathcal{G})$  and  $L(\mathcal{G})$  the set of nodes and the set of lines, respectively. A path between two nodes is the minimal subset of  $L(\mathcal{G})$  connecting the two nodes. A graph is planar if it can be drawn in a plane without graph lines crossing.

A *tree* is a planar graph  $\mathcal{G}$  containing no closed loops. Consider a tree  $\mathcal{G}$  with a single special node  $v_0$ : this introduces a natural partial ordering on the set of lines and nodes, and one can imagine that each line carries an arrow pointing toward the node  $v_0$ . We add an extra oriented line  $\ell_0$  exiting the special node  $v_0$ ; the added line will be called the *root line* and the point it enters (which is not a node) will be called the *root* of the tree. In this way we obtain a *tree*  $\theta$  defined by  $N(\theta) = N(\mathcal{G})$  and  $L(\theta) = L(\mathcal{G}) \cup \ell_0$ . A *labeled tree* is a rooted tree  $\theta$  together with a label function defined on the sets  $L(\theta)$  and  $N(\theta)$ .

We call *equivalent* two rooted trees which can be transformed into each other by continuously deforming the lines in the plane in such a way that the lines do not cross each other. We can extend the notion of equivalence also to labeled trees, by considering equivalent two labeled trees if they can be transformed into each other in such a way that the labels also match. In the following we shall deal mostly with nonequivalent labeled trees: for simplicity, where no confusion can arise, we call them just trees.

Given two nodes  $v, w \in N(\theta)$ , we say that  $w < v$  if  $v$  is on the path connecting  $w$  to the root line. We can identify a line  $\ell$  through the node  $v$  it exits by writing  $\ell = \ell_v$ .

We call *internal nodes* the nodes such that there is at least one line entering them, and *end-points* the nodes which have no entering line. We denote with  $L(\theta)$ ,  $V(\theta)$  and  $E(\theta)$  the set of lines, internal nodes and end-points, respectively. Of course  $N(\theta) = V(\theta) \cup E(\theta)$ .

The number of unlabeled trees with  $k$  nodes (and hence with  $k$  lines) is bounded by  $2^{2k}$ , which is a bound on the number of random walks with  $2k$  steps (Gentile and Mastropietro 2001).

For each node  $v$  denote by  $S(v)$  the set of the lines entering  $v$  and set  $s_v = |S(v)|$ . Hence  $s_v = 0$  if  $v$  is an end-node, and  $s_v \geq 1$  if  $v$  is an internal node. One has

$$\sum_{v \in N(\theta)} s_v = \sum_{v \in V(\theta)} s_v = k - 1; \quad (20)$$

this can be easily checked by induction on the order of the tree. An example of unlabeled tree is represented in Fig. 1.

For further details on graphs and trees we refer to the literature; cf. for instance Harary (1969).

### Labels and Diagrammatic Rule

We associate with each node  $v \in N(\theta)$  a *mode* label  $v_1 \in \mathbb{Z}^d$ , and with each line  $\ell \in L(\theta)$  a *momentum* label  $v_\ell \in \mathbb{Z}^d$ , with the constraint

$$v_{\ell_v} = \sum_{\substack{w \in N(\theta) \\ w < v}} v_w = v_v + \sum_{\ell \in S(v)} v_\ell, \quad (21)$$

which represents a *conservation rule* for each node.

Call  $\mathcal{T}_{k,v}$  the set of all trees  $\theta$  with  $k$  nodes and momentum  $v$  associated with the root line. We call  $k$  and  $v$  the order and the momentum of  $\theta$ , respectively.

We want to show that trees naturally arise when studying Eq. (19). Let  $h_v^{(k)}$  be represented with the graph element in Fig. 2 as a line with label  $v$  exiting from a ball with label  $(k)$ .

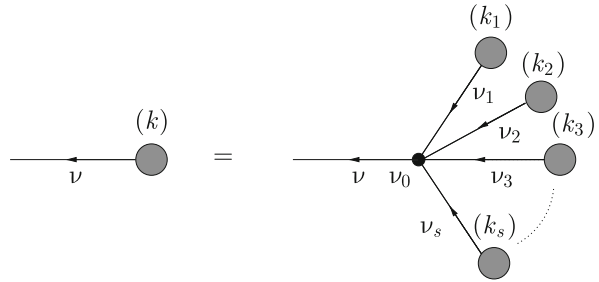
Then we can represent Eq. (19) graphically as depicted in Fig. 3. Simply represent each factor  $h_{v_i}^{(k_i)}$  on the right hand side as a graph element according to Fig. 2. The lines of all such graph



**Diagrammatic Methods in Classical Perturbation Theory, Fig. 2** Graph element

### Diagrammatic Methods in Classical Perturbation

**Theory, Fig. 3** Graphical representation of the recursive equations



elements enter the same node  $v_0$ . This is a graphical expedient to recall the conservation rule: the momentum  $v$  of the root line is the sum of the mode label  $v_0$  of the node  $v_0$  plus the sum of the momenta of the lines entering  $v_0$ .

The first few orders  $k \leq 4$  are as depicted in Fig. 4. For each node the conservation rule (21) holds: for instance for  $k = 2$  one has  $v = v_1 + v_2$ , for  $k = 3$  one has  $v = v_1 + v_{\ell_1}$  and  $v_{\ell_1} = v_2 + v_3$  in the first tree and  $v = v_1 + v_2 + v_3$  in the second tree, and so on. Moreover one has to sum over all possible choices of the labels  $v_i$ ,  $v \in N(\theta)$ , which sum up to  $v$ .

Given any tree  $\theta \in \mathcal{T}_{k,v}$  we associate with each node  $v \in N(\theta)$  a node factor  $F_v$  and with each line  $\ell \in L(\theta)$  a propagator  $g_\ell$ , by setting

$$F_v := \frac{1}{s_v!} (iv_v)^{s_v+1} f_{v_v}, \quad g_\ell := g(\omega \cdot v_\ell), \quad (22)$$

and define the *value* of the tree  $\theta$  as

$$\text{Val}(\theta) := \left( \prod_{v \in N(\theta)} F_v \right) \left( \prod_{\ell \in L(\theta)} g_\ell \right). \quad (23)$$

The propagators  $g_\ell$  are scalars, whereas each  $F_v$  is a tensor with  $s_v + 1$  indices, which can be associated with the  $s_v + 1$  lines entering or exiting  $v$ . In Eq. (23) the indices of the tensors  $F_v$  must be contracted: this means that if a node  $v$  is connected to a node  $v'$  by a line  $\ell$  then the indices of  $F_v$  and  $F_{v'}$  associated with  $\ell$  are equal to each other, and eventually one has to sum over all the indices except that associated with the root line.

For instance the value of the tree in Fig. 4 contributing to  $h_v^{(2)}$  is given by

$$\text{Val}(\theta) = (iv_1)^2 f_{v_1} (iv_2) f_{v_2} g(\omega \cdot v) g(\omega \cdot v_2),$$

with  $v_1 + v_2 = v$ , while the value of the last tree in Fig. 4 contributing to  $h_v^{(4)}$  is given by

$$\begin{aligned} \text{Val}(\theta) = & \frac{(iv_1)^4}{3!} f_{v_1} (iv_2) f_{v_2} (iv_3) f_{v_3} (iv_4) f_{v_4} \\ & \cdot g(\omega \cdot v) g(\omega \cdot v_2) g(\omega \cdot v_3) g(\omega \cdot v_4), \end{aligned}$$

with  $v_1 + v_2 + v_3 + v_4 = v$ .

It is straightforward to prove that one can write

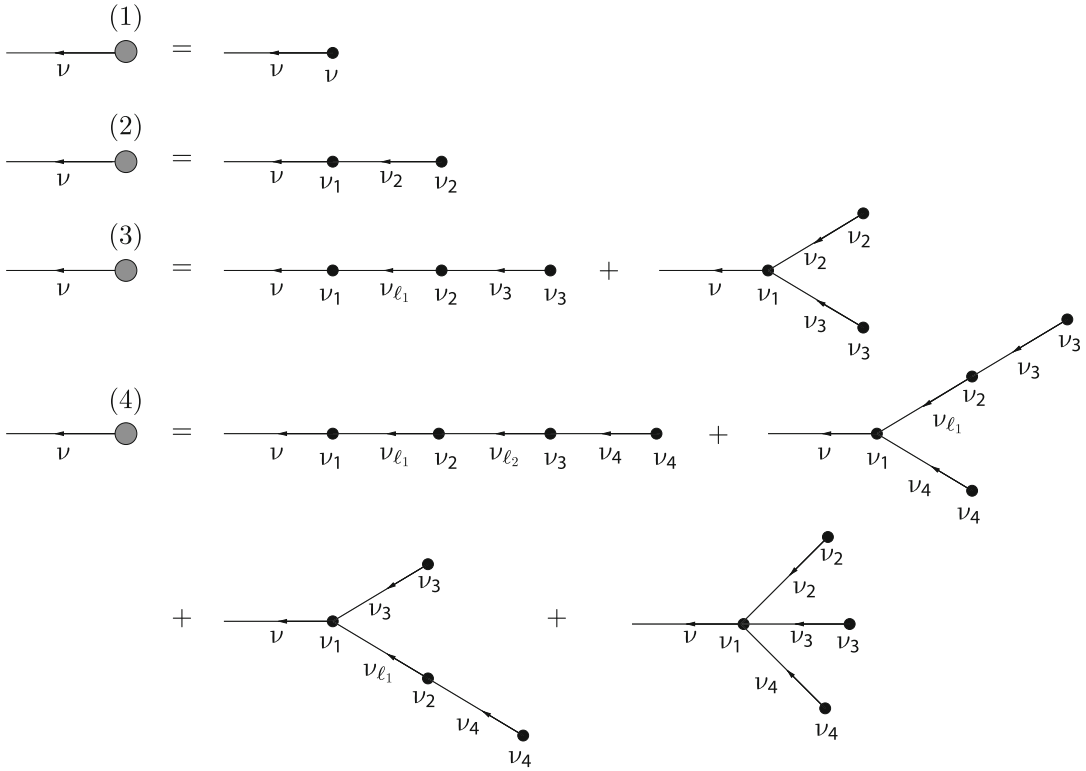
$$h_v^{(k)} = \sum_{\theta \in \mathcal{T}_{k,v}} \text{Val}(\theta), \quad v \neq 0, \quad k \geq 1. \quad (24)$$

This follows from the fact that the recursive equations (19) can be graphically represented through Fig. 3: one iterates the graphical representation of Fig. 3 until only graph elements of order  $k = 1$  appear, and if  $\theta$  is of order 1 (cf. Fig. 4) then  $\text{Val}(\theta) = (iv_v) f_{v_v} g(\omega \cdot v)$ .

Each line  $\ell \in L(\theta)$  can be seen as the root line of the tree consisting of all nodes and lines preceding  $\ell$ . The choice  $h_0^{(k)} = 0$  for all  $k \geq 1$  implies that no line can have zero momentum: in other words we have  $v_\ell \neq 0$  for all  $\ell \in L(\theta)$ .

Therefore in order to prove that Eq. (9) with  $h_v^{(k)}$  given by Eq. (24) solves formally, that is order by order, the Eq. (19), we have only to check that  $[\varepsilon \partial_x f(\omega t + h(\omega t))]_0^{(k)} = 0$  for all  $k \geq 1$ .

If we define  $g_\ell = 1$  for  $v_\ell = 0$ , then also the second relation in Eq. (19) can be graphically represented as in Fig. 3 by setting  $v = 0$  and requiring  $h_0^{(k)} = 0$ , which yields that the sum of the values of all trees on the right hand side must vanish. Note that this is not an equation to solve,



**Diagrammatic Methods in Classical Perturbation Theory, Fig. 4** Trees of lower orders

but just an identity that has to be checked to hold at all orders.

For instance for  $k = 2$  (the case  $k = 1$  is trivial) the identity  $[\varepsilon \partial_x f(\omega t + h(\omega t))]_0^{(2)} = 0$  reads (cf. the second line in Fig. 4)

$$\sum_{v_1+v_2=0} (iv_1)^2 f_{v_1}(iv_2) f_{v_2} g(\omega \cdot v_2) = 0,$$

which is found to be satisfied because the propagators are even in their arguments.

Such a cancellation can be graphically interpreted as follows. Consider the tree with mode labels  $v_1$  and  $v_2$ , with  $v_1 + v_2 = 0$ : its value is  $(iv_1)^2 f_{v_1}(iv_2) f_{v_2} g(\omega \cdot v_2)$ . One can detach the root line from the node with mode label  $v_1$  and attach it to the node with mode label  $v_2$ , and reverse the arrow of the other line so that it points toward the new root line. In this way we obtain a new tree (cf. Fig. 5): the value of the new tree is  $(iv_1) f_{v_1}(iv_2)^2 f_{v_2} g(\omega \cdot v_1)$ , where  $g(\omega \cdot v_1) = g(-\omega \cdot v_2) = g(\omega \cdot v_2)$ , so that the values of the

two trees contain a common factor  $(iv_1) f_{v_1}(iv_2) f_{v_2} g(\omega \cdot v_2)$  times an extra factor which is  $(iv_1)$  for the first tree and  $(iv_2)$  for the second tree. Hence the sum of the two values gives zero.

The cancellation mechanism described above can be generalized to all orders. Given a tree  $\theta$  one considers all trees which can be obtained by detaching the root line and attaching to the other nodes of the tree, and by reversing the arrows of the lines (when needed) to make them point toward the root line. Then one sums together the values of all the trees so obtained: such values contain a common factor times a factor  $iv_v$ , if  $v$  is the node which the root line exits (the only non-trivial part of the proof is to check that the combinatorial factors match each other: we refer to Gentile and Mastropietro (1996) for details). Hence the sum gives zero, as the sum of all the mode labels vanishes.

For instance for  $k = 3$  the cancellation operates by considering the three trees in Fig. 5: such trees



**Diagrammatic Methods in Classical Perturbation Theory, Fig. 5** Trees to be considered together to prove that  $[\varepsilon \partial f(\alpha)]_0^2 = 0$

can be considered to be obtained from each other by shifting the root line and consistently reversing the arrows of the lines.

In such a case the combinatorial factors of the node factors are different, because in the second tree the node factor associated with the node with mode label  $\nu_2$  contains a factor  $1/2$ : on the other hand if  $\nu_1 \neq \nu_3$  there are two nonequivalent trees with that shape (with the labels  $\nu_1$  and  $\nu_3$  exchanged between themselves), whereas if  $\nu_1 = \nu_3$  there is only one such tree, but then the first and third trees are equivalent, so that only one of them must be counted. Thus, by using that  $\nu_1 + \nu_2 + \nu_3 = 0$  – which implies  $g(\omega \cdot (\nu_2 + \nu_3)) = g(-\omega \cdot \nu_1)$  and  $g(\omega \cdot (\nu_1 + \nu_2)) = g(-\omega \cdot \nu_3)$  – in all cases we find that the sum of the values of the trees gives a common factor  $(i\nu_1)f_{\nu_1}(i\nu_2)^2f_{\nu_2}(i\nu_3)f_{\nu_3}g(\omega \cdot \nu_3)g(\omega \cdot \nu_1)$  times a factor 1 or  $1/2$  times  $i(\nu_1 + \nu_2 + \nu_3)$ , and hence vanishes: once more the property that  $g$  is even is crucial.

## Small Divisors

We want to study the convergence properties of the series

$$h(\omega t) = \sum_{v \in \mathbb{Z}^d} e^{i\omega \cdot vt} h_v, \quad h_v = \sum_{k=1}^{\infty} \varepsilon^k h_v^{(k)} \quad (25)$$

which has been shown to be well-defined as a formal power series for the models considered in section “Examples”.

Recall that the number of unlabeled trees of order  $k$  is bounded by  $2^{2k}$ . To sum over the labels we can confine ourselves to the mode labels, as the momenta are uniquely determined by the mode labels. If  $f$  is a trigonometric polynomial of degree  $N$ , that is  $f_v = 0$  for all  $v$  such that  $|v| := |v_1| + \dots + |v_d| > N$ , we have that  $h_v^{(k)} = 0$  for all  $|v| > kN$  (which can be easily proved by induction), and

moreover we can bound the sum over the mode labels of any tree of order  $k$  by  $(2N + 1)^{dk}$ . Finally we can bound

$$\prod_{v \in N(\theta)} |v_v|^{s_{v+1}} \leq \prod_{v \in N(\theta)} N^{s_v+1} \leq N^{2k}, \quad (26)$$

because of Eq. (20).

For the model (11), where  $g_\ell = 1$  in Eq. (22), we can bound

$$|h_v^{(k)}| \leq \sum_{v \in \mathbb{Z}^d} |h_v^{(k)}| \leq 2^{2k} (2N + 1)^{dk} N^{2k} \Phi^k, \quad \Phi = \max_{|v| \leq N} |f_v|, \quad (27)$$

which shows that the series (25) converges for  $\varepsilon$  small enough, more precisely for  $|\varepsilon| < \varepsilon_0$ , with

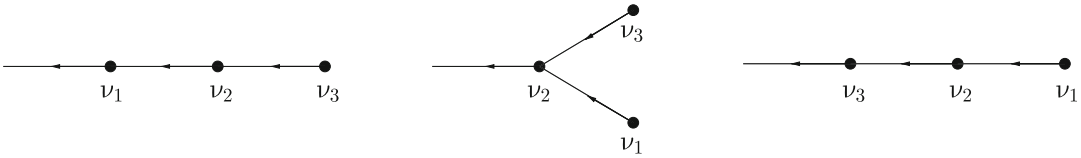
$$\varepsilon_0 := C_0 \left( 4N^2 \Phi (2N + 1)^d \right)^{-1}, \quad (28)$$

where  $C_0 = 1$ . Hence the function  $h(\omega t)$  in that case is analytic in  $\varepsilon$ . For  $d = 1$  and  $f(\alpha) = \cos \alpha$ , we can easily provide an exact expression for the coefficients  $h_v^{(k)}$ : all the computational difficulties reduce to a combinatorial check, which can be found in Gentile and van Erp (2005), and the formula (13) is recovered (Fig. 6).

However for the models where  $g_\ell \neq 1$ , the situation is much more involved: the propagators can be arbitrarily close to zero for  $v$  large enough. This is the so-called *small divisor problem*. The series (25) is formally well-defined, assuming only an irrationality condition on  $\omega$ . But to prove the convergence of the series, we need a stronger condition. For instance one can require the *standard Diophantine condition*

$$|\omega \cdot v| > \frac{\gamma}{|v|^\tau} \quad \forall v \neq 0, \quad (29)$$

for suitable positive constants  $\gamma$  and  $\tau$ . For fixed  $\tau > d - 1$ , the sets of vectors which satisfy Eq. (29) for some constant  $\gamma > 0$  has full Lebesgue measure in  $\mathbb{R}^d$  (Gallavotti 1983). We can also impose a weaker condition, known as the *Bryuno condition*, which can be expressed by requiring



**Diagrammatic Methods in Classical Perturbation Theory, Fig. 6** Trees to be considered together to prove that  $[\varepsilon \partial f(x)]_0^3 = 0$

$$\mathcal{B}(\omega) := \sum_{k=0}^{\infty} \frac{1}{2^k} \log \frac{1}{\min_{0 < |v| \leq 2^k} |\omega \cdot v|} < \infty. \quad (30)$$

$$\begin{cases} n \geq 1, & \text{if } 2^{-n}\gamma \leq |\omega \cdot v| < 2^{-(n-1)}\gamma, \\ n = 0, & \text{if } \gamma \leq |\omega \cdot v|, \end{cases} \quad (31)$$

We call *Diophantine vectors* and *Bryuno vectors* the vectors satisfying Eqs. (29) and (30), respectively. Note that any Diophantine vector satisfies Eq. (30).

If we assume only analyticity on  $f$  – that is, we remove the assumption that  $f$  be a trigonometric polynomial, – then we need a Diophantine condition, such as Eq. (29) or (30), also to show that the series (25) are well-defined as formal power series: the condition is needed, as one can easily check, in order to sum over themode labels.

In the case of the standard map  $|\omega \cdot v|$  must be replaced with  $\min_{p \in \mathbb{Z}} |\omega v - p|$ , and the function  $\mathcal{B}(\omega)$  can be expressed in terms of the best *approximants* of the number  $\omega$  (Davie 1994).

For the models with small divisors considered in section “Examples” convergence of the series (25) can be proved for  $\omega$  satisfying Eq. (29) or (30). But this requires more work, and, in particular, a detailed discussion of the product of propagators in Eq. (24). In the following we shall consider the case of Diophantine vectors, by referring to the bibliography for the Bryuno vectors (cf. section “Generalizations”).

## Multiscale Analysis

Consider explicitly the case  $g(\omega \cdot v) = (\omega \cdot v)^{-2}$  in Eq. (19) and  $\omega$  satisfying Eq. (29). We can introduce a label characterizing the size of the propagator: we say that  $v \in \mathbb{Z}^d$  is on *scale*

where  $\gamma$  is the constant appearing in Eq. (29), and we say that a line  $\ell$  has a scale label  $n_\ell = n$  if  $v_\ell$  is on scale  $n$ . If  $\mathfrak{N}_n(\theta)$  denotes the number of lines  $\ell \in L(\theta)$  with scale  $n_\ell = n$ , then we can bound in Eq. (23)

$$|\prod_{\ell \in L(\theta)} g_\ell| \leq \gamma^{-2k} \prod_{n=0}^{\infty} 2^{2n\mathfrak{N}_n(\theta)}, \quad (32)$$

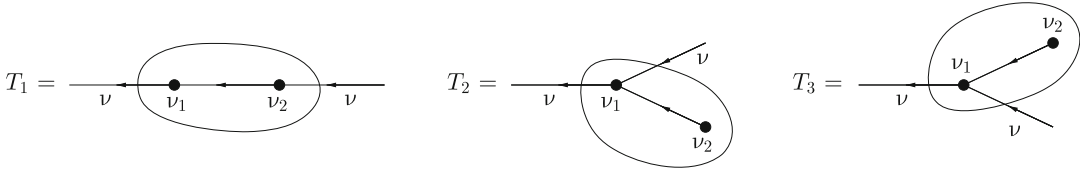
so that the problem is reduced to bounding  $\mathfrak{N}_n(\theta)$ .

The product of propagators gives problems when the small divisors “accumulate”. To make more precise the idea of accumulation we introduce the notion of cluster. Once all lines of a tree  $\theta$  have been given their scale labels, for any  $n \geq 0$  we can identify the maximal connected sets of lines with scale not larger than  $n$ . If at least one among such lines has scale equal to  $n$  we say that the set is a *cluster* on scale  $n$ . For instance consider the tree in Fig. 1, and assign the mode labels to the nodes: this uniquely fixes the momenta, and hence the scale labels, of the lines. Suppose that we have found the scales as in Fig. 7a. Then a cluster decomposition as in Fig. 7b follows. Given a cluster  $T$  call  $L(T)$  the set of lines of  $\theta$  contained in  $T$ , and denote by  $N(T)$  the set of nodes connected by such lines. We define  $k_T = |N(T)|$  the order of the cluster  $T$ .

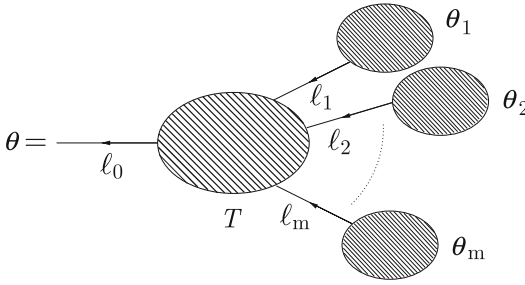
Any cluster has either one or no exiting line, and can have an arbitrary number of entering lines. We call *self-energy clusters* the clusters which have one exiting line and only one entering line and are such that both lines have the same momentum. This means that if  $T$  is a self-energy







**Diagrammatic Methods in Classical Perturbation Theory, Fig. 8** Self-energy clusters of order  $k = 2$



**Diagrammatic Methods in Classical Perturbation Theory, Fig. 9** Construction for the proof of the Siegel–Bryuno lemma

$m \geq 0$ , be the lines on scales  $\geq n$  which are the closest to  $\ell_0$ , and let  $\theta_1, \dots, \theta_m$  the trees with root lines  $\ell_1, \dots, \ell_m$ , respectively (cf. Fig. 9 – note that by construction all lines  $\ell$  in the subgraph  $T$  have scales  $n_\ell < n$ , so that if  $n_{\ell_0} \geq n$  then  $T$  is necessarily a cluster). If either  $\ell_0$  is not on scale  $n$  or it is on scale  $n$  but exits a self-energy cluster then  $\mathfrak{N}_n^*(\theta) = \mathfrak{N}_n^*(\theta_1) + \dots + \mathfrak{N}_n^*(\theta_m)$  and the bound  $\mathfrak{N}_n^*(\theta) \leq E(n, k)$  follows by the inductive hypothesis. If  $\ell_0$  does not exit a self-energy cluster and  $n_{\ell_0} = n$  then  $\mathfrak{N}_n^*(\theta) = 1 + \mathfrak{N}_n^*(\theta_1) + \dots + \mathfrak{N}_n^*(\theta_m)$ , and the lines  $\ell_1, \dots, \ell_m$  enter a cluster  $T$  with  $k_T = k - (k_1 + \dots + k_m)$ , where  $k_1, \dots, k_m$  are the orders of  $\theta_1, \dots, \theta_m$ , respectively. If  $m \geq 2$  the bound  $\mathfrak{N}_n^*(\theta) \leq E(n, k)$  follows once more by the inductive hypothesis. If  $m = 0$  then  $\mathfrak{N}_n^*(\theta) = 1$ ; on the other hand for  $\ell_0$  to be on scale  $n_{\ell_0} = n$  one must have  $|\omega \cdot v_{\ell_0}| < 2^{-n+1}\gamma$  (see Eq. 31), which, by the Diophantine condition (29), implies  $Nk \geq |v_{\ell_0}| > 2^{(n-1)/\tau}$ , hence  $E(n, k) > 1$ . If  $m = 1$  call  $v_1$  and  $v_2$  the momenta of the lines  $\ell_0$  and  $\ell_1$ , respectively. By construction  $T$  cannot be a self-energy cluster, hence  $v_1 \neq v_2$ , so that, by the Diophantine condition (29),

$$\begin{aligned} 2^{-n+2}\gamma &\geq |\omega \cdot v_1| + |\omega \cdot v_2| \geq |\omega \cdot (v_1 - v_2)| \\ &> \frac{\gamma}{|v_1 - v_2|^\tau}, \end{aligned} \quad (36)$$

because  $n_{\ell_0} = n$  and  $n_{\ell_1} \geq n$ . Thus, one has

$$Nk_T \geq \sum_{v \in N(T)} |v_v| \geq |v_1 - v_2| > 2^{(n-2)/\tau}, \quad (37)$$

hence  $T$  must contain “many nodes”. In particular, one finds also in this case  $\mathfrak{N}_n^*(\theta) = 1 + \mathfrak{N}_n^*(\theta_1) \leq 1 + E(n, k_1) \leq 1 + E(n, k) - E(n, k_T) \leq E(n, k)$ , where we have used that  $E(n, k_T) \geq 1$  by Eq. (37).

The argument above shows that small divisors can accumulate only by allowing self-energy clusters. That accumulation really occurs is shown by the example in Fig. 10, where a tree  $\theta$  of order  $k$  containing a chain of  $p$  self-energy clusters is depicted. Assume for simplicity that  $k/3$  is an integer: then if  $p = k/3$  the subtree  $\theta_1$  with root line  $\ell$  is of order  $k/3$ . If the line  $\ell$  entering the rightmost self-energy cluster  $T_p$  has momentum  $v$ , also the lines exiting the  $p$  self-energy clusters have the same momentum  $v$ . Suppose that  $|v| \approx Nk/3$  and  $|\omega \cdot v| \approx \gamma/|v|^\tau$  (this is certainly possible for some  $v$ ). Then the value of the tree  $\theta$  grows like  $a_1^k (k!)^{a_2}$ , for some constants  $a_1$  and  $a_2$ : a bound of this kind prevents the convergence of the perturbation series (25).

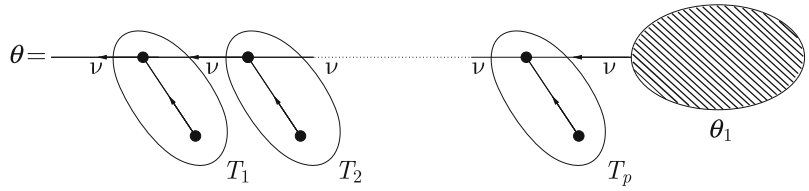
If no self-energy clusters could occur (so that  $\mathfrak{R}_n(\theta) = 0$ ) the Siegel–Bryuno lemma would allow us to bound in Eq. (32)

$$\begin{aligned} \prod_{n=0}^{\infty} 2^{2n_n(\theta)} &= \prod_{n=0}^{\infty} 2^{2n_n^*(\theta)} \leq \exp \left( C_1 k \sum_{n=0}^{\infty} n 2^{-n/\tau} \right) \\ &\leq C_2^k, \end{aligned} \quad (38)$$



### Diagrammatic Methods in Classical Perturbation Theory, Fig. 10

Example of accumulation of small divisors because of the self-energy clusters



for suitable constants  $C_1$  and  $C_2$ . In that case convergence of the series for  $|\varepsilon| < \varepsilon_0$  would follow, with  $\varepsilon_0$  defined as in Eq. (26) with  $C_0 = \gamma^2/C_2$ . However, there are self-energy clusters and they produce factorials, as the example in Fig. 10 shows, so that we have to deal with them.

### Resummation

Let us come back to the Eq. (10). If we expand  $g(\omega \cdot v)[\varepsilon \partial_{\mathcal{A}} f(\alpha)]_v^{(k)}$  in trees according to the diagrammatic rules described in section “Trees and Graphical Representation”, we can distinguish between contributions in which the root line exits a self-energy cluster  $T$ , that we can write as

$$g(\omega \cdot v) \sum_{T: k_T < k} \mathcal{V}_T(\omega \cdot v) h_v^{(k-k_T)}, \quad (39)$$

and all the other contributions, that we denote by  $[\varepsilon \partial_{\mathcal{A}} f(\alpha)]_v^{(k)*}$ . We can shift the contributions (39) to the left hand side of Eq. (10) and divide by  $g(\omega \cdot v)$ , so to obtain

$$\begin{aligned} \mathcal{D}(\omega \cdot v) h_v^{(k)} - \sum_{T: k_T = k} \mathcal{V}_T(\omega \cdot v) h_v^{(k-k_T)} \\ = [\varepsilon \partial_{\mathcal{A}} f(\omega t + h)]_v^{(k)*}. \end{aligned} \quad (40)$$

where  $\mathcal{D}(\omega \cdot v) = 1/g(\omega \cdot v) = (\omega \cdot v)^2$ . By summing over  $k$  and setting

$$M(\omega \cdot v; \varepsilon) = \sum_{k=1}^{\infty} \varepsilon^k \sum_{T: k_T = k} V_T(\omega \cdot v), \quad (41)$$

then Eq. (40) gives

$$\begin{aligned} \mathcal{D}(\omega \cdot v) h_v = [\varepsilon \partial_{\mathcal{A}} f(\omega t + h)]_v^*, \\ \mathcal{D}(\omega \cdot v) := \mathcal{D}(\omega \cdot v) - M(\omega \cdot v; \varepsilon) \end{aligned} \quad (42)$$

The motivation for proceeding in this way is that, at the price of changing the operator  $D$  into  $\mathcal{D}$ , hence of changing the propagators, lines exiting self-energy clusters no longer appear. Therefore, in the tree expansion of the right hand side of the equation, we have eliminated the self-energy clusters, that is the source of the problem of accumulation of small divisors.

Unfortunately the procedure described above has a problem:  $M(\omega \cdot v; \varepsilon)$  itself is a sum of self-energy clusters, which can still contain some other self-energy clusters on lower scales. So finding a good bound for  $M(\omega \cdot v; \varepsilon)$  could have the same problems as for the values of the trees.

To deal with such a difficulty we modify the prescription by proceeding recursively, in the following sense. Let us start from the momenta  $v$  which are on scale  $n = 0$ . Since there are no self-energy clusters with exiting line on scale  $n = 0$ , for such  $v$  one has  $M(\omega \cdot v; \varepsilon) = 0$ . Next, we consider the momenta  $v$  which are on scale  $n = 1$ , and we can write Eq. (42), where now all self-energy clusters  $T$  whose values contribute to  $M(\omega \cdot v; \varepsilon)$  cannot contain any self-energy clusters, because the lines  $\ell \in T$  are on scale  $n_\ell = 0$ . Then, we consider the momenta  $v$  which are on scale  $n = 2$ : again all the self-energy clusters contributing to  $M(\omega \cdot v; \varepsilon)$  do not contain any self-energy clusters, because the lines on scale  $n = 0, 1$  cannot exit self-energy clusters by the construction of the previous step, and so on. The conclusion is that we have obtained a different expansion for  $h(\omega t)$ , that we call a *resummed series*,

$$h(\omega t) = \sum_{v \in \mathbb{Z}^d} e^{i\omega \cdot vt} h_v, \quad h_v = \sum_{k=1}^{\infty} \varepsilon^k h_v^{[k]}(\varepsilon) \quad (43)$$

where the self-energy clusters do not appear any more in the tree expansion and the propagators

must be defined recursively: the propagator  $g_\ell$  of a line  $\ell$  on scale  $n_\ell = n$  and momentum  $v_\ell = v$  is the matrix

$$g_\ell := g^{[n]}(\omega \cdot v; \varepsilon) = \left( D(\omega \cdot v) - \mathcal{M}^{[n]}(\omega \cdot v; \varepsilon) \right)^{-1} \quad (44)$$

with

$$\mathcal{M}^{[n]}(\omega \cdot v; \varepsilon) := \sum_{T: n_T < n} \varepsilon^{k_T} \mathcal{V}_T(\omega \cdot v) \quad (45)$$

where the value  $\mathcal{V}_T(\omega \cdot v)$  is written in accord with Eq. (34), with all the lines  $\ell' \in L(T)$  on scales  $n_{\ell'} < n$  and the corresponding propagators  $g_{\ell'}$  expressed in terms of matrices  $\mathcal{M}^{[n_{\ell'}]}(\omega \cdot v_{\ell'}; \varepsilon)$  as in Eq. (44). By construction, the new propagators depend on  $\varepsilon$ , so that the coefficients  $h_v^{[k]}(\varepsilon)$  depend explicitly on  $\varepsilon$ : hence Eq. (43) is not a power series expansion.

The coefficients  $h_v^{[k]}(\varepsilon)$  still admit a tree expansion

$$h_v^{[k]}(\varepsilon) = \sum_{\theta \in \mathcal{T}_{k,v}^{\mathcal{R}}} \text{Val}(\theta), \quad v \neq 0, k \geq 1, \quad (46)$$

$$\text{Val}(\theta) := \left( \prod_{v \in N(\theta)} F_v \right) \left( \prod_{\ell \in L(\theta)} g^{[n_\ell]}(\omega \cdot v_\ell; \varepsilon) \right),$$

which replaces Eq. (24). In particular  $\mathcal{T}_{k,v}^{\mathcal{R}}$  is defined as the set of renormalized trees of order  $k$  and momentum  $v$ , where “renormalized” means that the trees do not contain any self-energy clusters. Now the propagators are matrices: the contractions of the indices in Eq. (46) yields that if  $\ell$  connects  $v$  to  $v'$  the indices of  $F_v$  and  $F_{v'}$  associated with  $\ell$  must be equal to the column and row indices of  $g_\ell$ , respectively.

Since for any tree  $\theta \in \mathcal{T}_{k,v}^{\mathcal{R}}$  one has  $\mathfrak{N}_n(\theta) = \mathfrak{N}_n^*(\theta)$ , we can bound the product of propagators according to Eq. (36) *provided the propagators on scale  $n$  can still be bounded proportionally to  $2^{2n}$* . This is certainly not obvious, because of the extra term  $\mathcal{M}^{[n]}(\omega \cdot v; \varepsilon)$  appearing in Eq. (44).

It is a remarkable cancellation that  $\mathcal{M}^{[n]}(x; \varepsilon)$  vanishes up to second order, that is  $\mathcal{M}^{[n]}(x; \varepsilon) = O(x^2)$ , so that, by taking into account also that  $\mathcal{V}_T(x) \neq 0$  requires  $k_T \geq 2$ , we can write, for some constant  $C$ ,

$$\mathcal{M}^{[n]}(x; \varepsilon) = \varepsilon^2 x^2 \overline{\mathcal{M}}^{[n]}(x; \varepsilon), \quad \left\| \overline{\mathcal{M}}^{[n]}(x; \varepsilon) \right\| \leq C, \quad (47)$$

where  $\| \cdot \|$  denotes – say – the uniform norm. The cancellation leading to Eq. (47) can be proved as discussed in section “[Multiscale Analysis](#)” – where it has been explicitly proved for  $k = 2$  in the absence of resummation. Now the propagators are matrices, but one can prove (by induction on  $n$ ) that  $\mathcal{M}^{[n]}(x; \varepsilon) = (\mathcal{M}^{[n]}(-x; \varepsilon))^T = (\mathcal{M}^{[n]}(x; \varepsilon))^\dagger$ , with  $T$  and  $\dagger$  denoting transposition and adjointness, respectively, and this is enough to see that the same cancellation mechanism applies (Gallavotti et al. 2004).

Thus Eq. (47) implies that

$$\|g^{[n]}(x; \varepsilon)\| = \left\| \left( x^2 - \varepsilon^2 x^2 \overline{\mathcal{M}}^{[n]}(x; \varepsilon) \right)^{-1} \right\| \leq \frac{2}{x^2}, \quad (48)$$

and the same argument as used in section “[Small Divisors](#)” implies that the series in Eq. (43) for  $h_v$  converges for  $|\varepsilon| < \varepsilon_0$ , where  $\varepsilon_0$  is defined as in Eq. (26) with  $C_0 = 2\gamma^2/C_2$ . Moreover Eq. (47) also implies that  $h(\omega t)$  is analytic in  $\varepsilon$  (notwithstanding that the expansion is not a power expansion), so that we can say a posteriori that the original power series (25) also converges.

## Generalizations

The case where the function  $G(u)$  in Eq. (1) does not vanish is notationally more involved, and it is explicitly discussed in Gentile (2006a).

Extensions of the results described above to Hamiltonian functions more general than Eq. (5) require only some notational complications, and can be found in Gentile and Mastropietro (1996) for anisochronous systems and in Bartuccelli and

Gentile (2002) for isochronous systems. The case (5) with  $f$  a real analytic function was first discussed by using trees in Chierchia and Falcolini (1994).

### Lower-Dimensional Tori

A tree formalism for hyperbolic lower-dimensional tori was introduced and studied in Gallavotti (1994b), Gentile (1995) and Gallavotti, Gentile, and Mastropietro (1999), for systems consisting of a set of rotators interacting with a pendulum. In that case also the stable and unstable manifolds (*whiskers*) of the tori were studied with diagrammatic techniques.

For Hamiltonian functions of the form (5) solutions of the form (43) describe quasi-periodic motions with frequency vector  $\omega$  on a  $d$ -dimensional manifold. If we fix  $t$ , say  $t = 0$ , and keep  $\alpha_0$  as a parameter, we obtain a parametrization of the manifold in terms of  $\alpha_0 \in \mathbb{T}^d$ , hence the manifold is an invariant torus. We say that the torus is a maximal torus.

We can study the problem of persistence of lower-dimensional tori by considering unperturbed solutions  $\alpha(t) = \alpha_0 + \omega t$ , where the components of  $\omega$  are rationally dependent. For instance we can imagine that there exist  $s$  linearly independent vectors  $\widehat{v}_1, \dots, \widehat{v}_s \in \mathbb{Z}^d$  such that  $\omega \cdot \widehat{v}_k = 0$  for  $k = 1, \dots, s$ . In that case, we say that the unperturbed torus is a resonant torus of order  $s$ . We can imagine performing a linear change of variables which transforms the frequency vector into a new vector, that we still denote with  $\omega$ , such that  $\omega = (\overline{\omega}, 0)$ , where  $\overline{\omega} \in \mathbb{R}^r$  and  $0 \in \mathbb{R}^s$ , with  $r + s = d$ , and  $\overline{\omega}$  is irrational. This naturally suggests that we write  $\alpha = (\overline{\alpha}, \beta)$ , with  $\overline{\alpha} \in \mathbb{T}^r$  and  $\beta \in \mathbb{T}^s$ . More generally, here and henceforth in this subsection for any vector  $v \in \mathbb{R}^d$  we denote by  $\overline{v}$  the vector in  $\mathbb{R}^r$  whose components are the first  $r$  components of  $v$ . For instance for the initial phase  $\alpha_0$  we write  $\alpha_0 = (\overline{\alpha}_0, \beta_0)$ .

In general the resonant torus is destroyed by the perturbation, and only some lower-dimensional tori persist under perturbation. We assume on  $\overline{\omega}$  one of the Diophantine conditions of section “Small Divisors” in  $\mathbb{R}^r$ , for instance  $|\overline{\omega} \cdot \overline{v}| > \gamma/|\overline{v}|^\tau$  for all  $\overline{v} \in \mathbb{Z}^r, \overline{v} \neq 0$ .

To prove the existence of a maximal torus for the system with Hamiltonian function (5) we needed no condition on the perturbation  $f$ . On the contrary to prove existence of lower-dimensional tori, we need some *non-degeneracy condition*: by defining

$$f_0(\beta) = \int_{\mathbb{T}^r} \frac{d\overline{\alpha}}{(2\pi)^r} f(\overline{\alpha}, \beta), \quad (49)$$

if  $\partial_\beta f_0(\beta_*) = 0$  for some  $\beta_*$  then we assume that the matrix  $\partial_\beta^2 f_0(\beta_*)$  is positive definite (more generally one could assume it to be non-singular, that is  $\det \partial_\beta^2 f_0(\beta_*) \neq 0$ ).

The formal analysis can be carried out as in section “Trees and Graphical Representation”, with the only difference being that now the compatibility conditions  $[\varepsilon \partial_\alpha f(\alpha)]_v^{(k)} = 0$  have to be imposed for all  $v$  such that  $\overline{v} = 0$ , because  $\omega \cdot v = \overline{\omega} \cdot \overline{v}$ . It turns out to be an identity only for the first  $r$  components (this is trivial for  $k = 1$ , whereas it requires some work for  $k > 1$ ). For the last  $s$  components, for  $k = 1$  it reads  $\partial_\beta f_0(\beta_0) = 0$ , hence it fixes  $\beta_0$  to be a stationary point for  $f_0(\beta)$ , while for higher values of  $k$  it fixes the corrections of higher order of these values (to do this we need the non-degeneracy condition). Thus, we are free to choose only  $\overline{\alpha}_0$  as a free parameter, since the last  $s$  components of  $\alpha_0$  have to be fixed.

Clusters and self-energy clusters are defined as in section “Multiscale Analysis”. Note that only the first  $r$  components  $\overline{v}$  of the momenta  $v$  intervene in the definition of the scales – again because  $\omega \cdot v = \overline{\omega} \cdot \overline{v}$ . In particular, in the definition of self-energy clusters, in Eq. (33) we must replace  $v_v$  with  $\overline{v}_v$ . Thus, already to first order the value of a self-energy cluster can be non-zero: for  $k_T = 1$ , that is for  $T$  containing only a node  $v$  with mode label  $(\overline{v}_v, \tilde{v}_v) = (0, \tilde{v}_v)$ , the matrix  $\mathcal{V}_T(x; \varepsilon)$  is of the form

$$\mathcal{V}_T(x) = \begin{pmatrix} 0 & 0 \\ 0 & b_{\overline{v}_v} \end{pmatrix},$$

$$(b_{\overline{v}_v})_{i,j} = e^{i\tilde{v}_v \cdot \beta_0} (i\tilde{v}_{v,i}) (i\tilde{v}_{v,j}) f(0, \widehat{v}_v), \quad (50)$$

with  $i, j = r + 1, \dots, d$ . If we sum over  $\tilde{v}_v \in \mathbb{Z}^s$  and multiply times  $\varepsilon$ , we obtain

$$M_0 := \begin{pmatrix} 0 & 0 \\ 0 & \varepsilon B \end{pmatrix},$$

$$B_{ij} = \sum_{\tilde{v} \in \mathbb{Z}^s} e^{i\tilde{v} \cdot \beta_0} (i\tilde{v}_i) (i\tilde{v}_j) f_{(0, \tilde{v})} = \partial_{\beta_i} \partial_{\beta_j} f_0(\beta_0). \quad (51)$$

The  $s \times s$  block  $B$  is non-zero: in fact, the non-degeneracy condition yields that it is invertible.

To higher orders one finds that the matrix  $\mathcal{M}^{[n]}(x; \varepsilon)$ , with  $x = \omega \cdot v = \overline{\omega} \cdot \bar{v}$ , is a self-adjoint matrix and  $\mathcal{M}^{[n]}(x; \varepsilon) = (\mathcal{M}^{[n]}(-x; \varepsilon))^T$ , as in the case of maximal tori. Moreover the corresponding eigenvalues  $\lambda_i^{[n]}(x; \varepsilon)$  satisfy  $\lambda_i^{[n]}(x; \varepsilon) = O(\varepsilon^2 x^2)$  for  $i = 1, \dots, r$  and  $\lambda_i^{[n]}(x; \varepsilon) = O(\varepsilon)$  for  $i = r + 1, \dots, d$ ; this property is not trivial because of the off-diagonal blocks (which in general do not vanish at orders  $k \geq 2$ ), and to prove it one has to use the self-adjointness of the matrix  $\mathcal{M}^{[n]}(x; \varepsilon)$ . More precisely one has  $\lambda_i^{[n]}(x; \varepsilon) = \varepsilon a_i + O(\varepsilon^2)$  for  $i > r$ , where  $a_r + 1, \dots, a_d$  are the  $s$  eigenvalues of the matrix  $B$  in Eq. (51). From this point on the discussion proceeds in a very different way according to the sign of  $\varepsilon$  (recall that we are assuming that  $a_i > 0$  for all  $i > r$ ).

For  $\varepsilon < 0$  one has  $\lambda_i^{[n]}(x; \varepsilon) = a_i \varepsilon + O(\varepsilon^2) < 0$  for  $i > r$ , so that we can bound the last  $s$  eigenvalues of  $x^2 - \mathcal{M}^{[n]}(x; \varepsilon)$  with  $x^2$ , and the first  $r$  with  $x^2/2$  by the same argument as in section “Resummation”. Hence we obtain easily the convergence of the series (43); of course, analyticity at the origin is prevented because of the condition  $\varepsilon < 0$ . We say in that case that the lower-dimensional tori are *hyperbolic*. We refer to Gallavotti and Gentile (2002) and Gallavotti et al. (2004) for details.

The case of *elliptic* lower-dimensional tori – that is  $\varepsilon > 0$  when  $a_i > 0$  for all  $i > r$  – is more difficult. Essentially the idea is as follows (we only sketch the strategy: the details can be found in Gentile and Gallavotti (2005)). One has to define the scales recursively, by using a variant, first introduced in Gentile (2003), of the

resummation technique described in section “Resummation”. We say that  $v$  is on scale 0 if  $|\overline{\omega} \cdot \bar{v}| \geq \gamma$  and on scale  $[\geq 1]$  otherwise: for  $v$  on scale 0 we write Eq. (42) with  $M = M_0$ , as given in Eq. (51). This defines the propagators of the lines  $\ell$  on scale  $n = 0$  as

$$g_\ell = g^{[0]}(\omega \cdot v_\ell) = \left( (\overline{\omega} \cdot \bar{v}_\ell)^2 - M_0 \right)^{-1}. \quad (52)$$

Denote by  $\lambda_i$  the eigenvalues of  $M_0$ : given  $v$  on scale  $[\geq 1]$  we say that  $v$  is on scale 1 if  $2^{-1}\gamma \leq \min_{i=1, \dots, d} \sqrt{|\overline{\omega} \cdot \bar{v}|^2 - \lambda_i}$ , and on scale  $[\geq 2]$  if  $\min_{i=1, \dots, d} \sqrt{|\overline{\omega} \cdot \bar{v}|^2 - \lambda_i} < 2^{-1}\gamma$ . For  $v$  on scale 1 we write Eq. (42) with  $M$  replaced by  $\mathcal{M}^{[0]}(\omega \cdot \bar{v}; \varepsilon)$ , which is given by  $M_0$  plus the sum of the values of all self-energy clusters  $T$  on scale  $n_T = 0$ . Then the propagators of the lines  $\ell$  on scale  $n_\ell = 1$  is defined as

$$g_\ell = g^{[1]}(\omega \cdot v_\ell) = \left( (\overline{\omega} \cdot \bar{v}_\ell)^2 - \mathcal{M}^{[0]}(\overline{\omega} \cdot \bar{v}_\ell; \varepsilon) \right)^{-1}. \quad (53)$$

Call  $\lambda_i^{[n]}(x; \varepsilon)$  the eigenvalues of  $\mathcal{M}^{[n]}(x; \varepsilon)$ : given  $v$  on scale  $[\geq 2]$  we say that  $v$  is on scale 2 if  $2^{-2}\gamma \leq \min_{i=1, \dots, d} \sqrt{|\overline{\omega} \cdot \bar{v}|^2 - \lambda_i^{[0]}(\overline{\omega} \cdot \bar{v}; \varepsilon)}$ , and on scale  $[\geq 3]$  if  $\min_{i=1, \dots, d} \sqrt{|\overline{\omega} \cdot \bar{v}|^2 - \lambda_i^{[0]}(\overline{\omega} \cdot \bar{v}; \varepsilon)} < 2^{-2}\gamma$ . For  $v$  on scale 2 we write Eq. (42) with  $M$  replaced by  $\mathcal{M}^{[1]}(\overline{\omega} \cdot \bar{v}; \varepsilon)$ , which is given by  $\mathcal{M}^{[0]}(\overline{\omega} \cdot \bar{v}; \varepsilon)$  plus the sum of the values of all self-energy clusters  $T$  on scale  $n_T = 1$ . Thus, the propagators of the lines  $\ell$  on scale  $n_\ell = 2$  will be defined as

$$g_\ell = g^{[2]}(\omega \cdot v_\ell) = \left( (\overline{\omega} \cdot \bar{v}_\ell)^2 - \mathcal{M}^{[1]}(\overline{\omega} \cdot \bar{v}_\ell; \varepsilon) \right)^{-1}, \quad (54)$$

and so on. The propagators are self-adjoint matrices, hence their norms can be bounded through the corresponding eigenvalues. In order to proceed as in sections “Multiscale Analysis” and “Resummation” we need some Diophantine conditions on these eigenvalues. We can assume for some  $\tau' > \tau$

$$\left| \overline{\omega} \cdot \bar{v} - \sqrt{\left| \lambda_i^{[n]}(\overline{\omega} \cdot \bar{v}; \varepsilon) \right|} \right| > \frac{\gamma}{|\bar{v}|^{\tau'}} \quad \forall \bar{v} \neq 0, \quad (55)$$

for all  $i = 1, \dots, d$  and  $n \geq 0$ . These are known as the first *Melnikov conditions*.

Unfortunately, things do not proceed so plainly. In order to prove a bound like Eq. (35), possibly with a different  $\tau'$  replacing  $\tau$ , we need to compare the propagators of the lines entering and exiting clusters  $T$  which are not self-energy clusters. This requires replacing Eq. (36) with

$$\begin{aligned} 2^{-n+2}\gamma &\geq \\ \left| \overline{\omega} \cdot (\bar{v}_1 - \bar{v}_2) \pm \sqrt{\left| \lambda_i^{[n]}(\overline{\omega} \cdot \bar{v}_1; \varepsilon) \right|} \pm \sqrt{\left| \lambda_j^{[n]}(\overline{\omega} \cdot \bar{v}_2; \varepsilon) \right|} \right| \\ &> \frac{\gamma}{|\bar{v}_1 - \bar{v}_2|^{\tau'}}, \end{aligned} \quad (56)$$

for all  $i, j = 1, \dots, d$  and choices of the signs  $\pm$ , and hence introduces further Diophantine conditions, known as the *second Melnikov conditions*.

The conditions in Eq. (56) turn out to be too many, because for all  $n \geq 0$  and all  $\bar{v} \in \mathbb{Z}^r$  such that  $\bar{v} = \bar{v}_1 - \bar{v}_2$  there are infinitely many conditions to be considered, one per pair  $(\bar{v}_1, \bar{v}_2)$ . However we can impose both the conditions (55) and (56) not for the eigenvalues  $\lambda_i^{[n]}(\overline{\omega} \cdot \bar{v}; \varepsilon)$ , but for some quantities  $\underline{\lambda}_i^{[n]}(\varepsilon)$  independent of  $\bar{v}$  and then use the smoothness of the eigenvalues in  $x$  to control  $(\overline{\omega} \cdot \bar{v})^2 - \lambda_i^{[n]}(\overline{\omega} \cdot \bar{v}; \varepsilon)$  in terms of  $(\overline{\omega} \cdot \bar{v})^2 - \underline{\lambda}_i^{[n]}(\varepsilon)$ . Eventually, beside the Diophantine condition on  $\overline{\omega}$ , we have to impose the Melnikov conditions

$$\begin{aligned} \left| \overline{\omega} \cdot \bar{v} - \sqrt{\left| \underline{\lambda}_i^{[n]}(\varepsilon) \right|} \right| &> \frac{\gamma}{|\bar{v}|^{\tau'}}, \\ \left| \overline{\omega} \cdot \bar{v} \pm \sqrt{\left| \underline{\lambda}_i^{[n]}(\varepsilon) \right|} \pm \sqrt{\left| \underline{\lambda}_j^{[n]}(\varepsilon) \right|} \right| &> \frac{\gamma}{|\bar{v}|^{\tau'}}, \end{aligned} \quad (57)$$

for all  $\bar{v} \neq 0$  and all  $n \geq 0$ . Each condition in Eq. (57) leads us to eliminate a small interval of

values of  $\varepsilon$ . For the values of  $\varepsilon$  which are left define  $h(\omega t)$  according to Eqs. (43) and (46), with the new definition of the propagators. If  $\varepsilon$  is small enough, say  $|\varepsilon| < \varepsilon_0$ , then the series (43) converges. Denote by  $\mathfrak{E} \subset [0, \varepsilon_0]$  the set of values of  $\varepsilon$  for which the conditions (57) are satisfied. One can prove that  $\mathfrak{E}$  is a *Cantor set*, that is a perfect, nowhere dense set. Moreover  $\mathfrak{E}$  has large relative Lebesgue measure, in the sense that

$$\lim_{\varepsilon \rightarrow 0} \frac{\text{meas}(\mathfrak{E} \cap [0, \varepsilon])}{\varepsilon} = 1, \quad (58)$$

provided  $\tau'$  in Eq. (57) is large enough with respect to  $\tau$ . The property (58) yields that, notwithstanding that we are eliminating infinitely many intervals, the measure of the union of all these intervals is small.

If  $a_i < 0$  for all  $i > r$  we reason in the same way, simply exchanging the role of positive and negative  $\varepsilon$ . On the contrary if  $a_i = 0$  for some  $i > r$ , the problem becomes much more difficult. For instance if  $s = 1$  and  $a_{r+1} = 0$ , then in general perturbation theory in  $\varepsilon$  is not possible, not even at a formal level. However, under some conditions, one can still construct fractional series in  $\varepsilon$ , and prove that the series can be resummed (Gallavotti et al. 2006).

### Other Ordinary Differential Equations

The formalism described above extends to other models, such as skew-product systems (Gentile 2006b) and systems with strong damping in the presence of a quasi-periodic forcing term (Gentile et al. 2005a).

As an example of skew-product system one can consider the linear differential equation  $\dot{x} = (\lambda A + \varepsilon f(\omega t))x$  on  $\text{SL}(2, \mathbb{R})$ , where  $\lambda \in \mathbb{R}$ ,  $\varepsilon$  is a small real parameter,  $\omega \in \mathbb{R}^n$  is an irrational vector, and  $A, f \in \mathfrak{sl}(2, \mathbb{R})$ , with  $A$  is a constant matrix and  $f$  an analytic function periodic in its arguments. Trees for skew-products were considered by Iserles and Nørsett (1999), but they used expansions in time, hence not suited for the study of global properties, such as quasi-periodicity.

Quasi-periodically forced one-dimensional systems with strong damping are described by

the ordinary differential equations  $\ddot{x} + \gamma\dot{x} + g(x) = f(\omega t)$ , where  $x \in \mathbb{R}$ ,  $\varepsilon = 1/\gamma$  is a small real parameter,  $\omega \in \mathbb{R}^n$  is irrational, and  $f, g$  are analytic functions ( $g$  is the “force”), with  $f$  periodic in its arguments.

We refer to the bibliography for details and results on the existence of quasi-periodic solutions.

### Bryuno Vectors

The diagrammatic methods can be used to prove the any unperturbed maximal torus with frequency vector which is a Bryuno vector persists under perturbation for  $\varepsilon$  small enough (Gentile 2007). One could speculate whether the Bryuno condition (30) is optimal. In general the problem is open. However, in the case of the standard map – see section “The Standard Map”, – one can prove (Berretti and Gentile 2001a; Davie 1994) that, by considering the radius of convergence  $\varepsilon_0$  of the perturbation series as a function of  $\omega$ , say  $\varepsilon_0 = \rho_0(\omega)$ , then there exists a universal constant  $C$  such that

$$|\log \rho_0(\omega) + 2\mathcal{B}(\omega)| \leq C. \quad (59)$$

In particular this yields that the invariant curve with rotation number  $\omega$  persists under perturbation if and only if  $\omega$  is a Bryuno number. The proof of Eq. (59) requires the study of a more refined cancellation than that discussed in section “Resummation”. We refer to Berretti and Gentile (2001a) and Gentile (2006a) for details.

Extensions to Bryuno vectors for lower-dimensional tori can also be found in Gentile (2007). For the models considered in section “Other Ordinary Differential Equations” we refer to Gentile (2006b) and Gentile et al. (2006).

### Partial Differential Equations

Existence of quasi-periodic solutions in systems described by one-dimensional nonlinear partial differential equations (finite-dimensional tori in infinite-dimensional systems) was first studied by Kuksin (1993), Craig and Wayne (1993), and Bourgain (1998). In these systems, even the case of periodic solutions yields small divisors, and hence requires a multiscale analysis. The study

of persistence of periodic solutions for nonlinear Schrödinger equations and nonlinear wave equations, with the techniques discussed here, can be found in Gentile and Mastropietro (2004), Gentile et al. (2005c), and Gentile and Procesi (2006).

The models are still described by Eq. (1), with  $G(u) = 0$ , but now  $D$  is given by  $D = \partial_t^2 - \Delta + \mu$  in the case of the wave equation and by  $D = i\partial_t - \Delta + \mu$  in the case of the Schrödinger equation, where  $\Delta$  is the Laplacian and  $\mu \in \mathbb{R}$ . In dimension 1, one has  $\Delta = \partial_x^2$ . If we look for periodic solutions with frequency  $\omega$  it can be convenient to pass to Fourier space, where the operator  $D$  acts as

$$D : e^{i\omega t + imx} \rightarrow (-\omega^2 n^2 + m^2 + \mu) e^{i\omega t + imx}, \quad (60)$$

for the wave equation; a similar expression holds for the Schrödinger equation. Therefore the kernel of  $D$  can be arbitrarily close to zero for  $n$  and  $m$  large enough.

Then one can consider, say, Eq. (1) for  $x \in [0, \pi]$  and  $F(u) = u^3$ , with Dirichlet boundary conditions  $u(0) = u(\pi) = 0$ , and study the existence of periodic solutions with frequency  $\omega$  close to some of the unperturbed frequencies. We refer to the cited bibliography for results and proofs.

## Conclusions and Future Directions

The diagrammatic techniques described above have been applied also in cases where no small divisors appear; cf. Berretti and Gentile (1999) and Gentile et al. (2007). Of course, such problems are much easier from a technical point of view, and can be considered as propaedeutic examples to become familiar with the treeformalism. Also the study of lower-dimensional tori becomes easy for  $r = 1$  (periodic solutions): in that case one has  $|\overline{\omega} \cdot \overline{v}| \geq |\overline{\omega}|$  for all  $\overline{v} \neq 0$ , so that the product of the propagators is bounded by  $|\overline{\omega}|^{-2k}$ , and one can proceed as in section “Small Divisors” to obtain analyticity of the solutions.

In the case of hyperbolic lower-dimensional tori, if  $\omega$  is a two-dimensional Diophantine vector of constant type (that is, with  $\tau = 1$ ) the conjugation function  $h$  can be proved to be Borel



summable (Costin et al. 2007). Analogous considerations hold for the one-dimensional systems in the presence of friction and of a quasiperiodic forcing term described in section “Other Ordinary Differential Equations”; in that case one has Borel summability also for one-dimensional  $\omega$ , that is for periodic forcing (Gentile et al. 2006). It would be interesting to investigate whether Borel summability could be obtained for higher values of  $\tau$ .

Recent existence of finite-dimensional tori in the nonlinear Schrödinger equation in higher dimensions was proved by Bourgain (2005). It would be interesting to investigate how far the diagrammatic techniques extend to deal with such higher-dimensional generalizations. The main problem is that (the analogues of) the second Melnikov conditions in Eq. (57) cannot be imposed.

In certain cases the tree formalism was extended to non-analytic systems, such as some quasi-integrable systems of the form (5) with  $f$  in a class of  $C^p$  functions for some finite  $p$  (Bonetto et al. 1998a, b). However, up to exceptional cases, the method described here seems to be intrinsically suited in cases in which the vector fields are analytic. The reason is that in order to exploit the expansion (3), we need that  $F$  be infinitely many times differentiable and we need a bound on the derivatives. It is a remarkable property that the perturbation series can be given a meaning also in cases where the solutions are not analytic in  $\varepsilon$ .

An advantage of the diagrammatic method is that it allows rather detailed information about the solutions, hence it could be more convenient than other techniques to study problems where the underlying structure is not known or too poor to exploit general abstract arguments.

Another advantage is the following. If one is interested not only in proving the existence of the solutions, but also in explicitly constructing them with any prefixed precision, this requires performing analytical or numerical computations with arbitrarily high accuracy. Then high perturbation orders have to be reached, and the easiest and most direct way to proceed is just through perturbation theory: so the approach illustrated here allows a unified treatment for both theoretical investigations and computational ones.

The resummation technique described in section “Resummation” can also be used for computational purposes. With respect to the naive power series expansion it can reduce the computation time required to approximate the solution within a prefixed precision. It can also provide accurate information on the analyticity properties of the solution. For instance, for the Kepler equation, Levi-Civita at the beginning of the last century described a resummation rule (see Levi-Civita 1954), which gives immediately the radius of convergence of the perturbation series. Of course, in the case of small divisor problems, everything becomes much more complicated.

## Bibliography

### Primary Literature

- Arnold VI (1963) Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian (Russian). *Uspehi Mat Nauk* 18(5):13–40
- Arnold VI, Kozlov VV, Neishtadt AI (1988) Dynamical systems III. In: *Encyclopaedia of mathematical sciences*, vol 3. Springer, Berlin
- Bartuccelli MV, Gentile G (2002) Lindstedt series for perturbations of isochronous systems: a review of the general theory. *Rev Math Phys* 14(2):121–171
- Berretti A, Gentile G (1999) Scaling properties for the radius of convergence of a Lindstedt series: the standard map. *J Math Pures Appl* 78(2):159–176
- Berretti A, Gentile G (2000) Scaling properties for the radius of convergence of a Lindstedt series: generalized standard maps. *J Math Pures Appl* 79(7):691–713
- Berretti A, Gentile G (2001a) Bryuno function and the standard map. *Commun Math Phys* 220(3):623–656
- Bonetto F, Gallavotti G, Gentile G, Mastropietro V (1998a) Quasi linear flows on tori: regularity of their linearization. *Commun Math Phys* 192(3):707–736
- Bonetto F, Gallavotti G, Gentile G, Mastropietro V (1998b) Lindstedt series, ultraviolet divergences and Moser’s theorem. *Ann Scuola Norm Sup Pisa Cl Sci* 26(3):545–593
- Bourgain J (1998) Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. *Ann Math* 148(2):363–439
- Bourgain J (2005) Green’s function estimates for lattice Schrödinger operators and applications. In: *Annals of mathematics studies* 158. Princeton University Press, Princeton
- Bricmont J, Gawędzki K, Kupiainen A (1999) KAM theorem and quantum field theory. *Commun Math Phys* 201(3):699–727



- Chierchia L, Falcolini C (1994) A direct proof of a theorem by Kolmogorov in Hamiltonian systems. *Ann Scuola Norm Sup Pisa Cl Sci* 21(4):541–593
- Costin O, Gallavotti G, Gentile G, Giuliani A (2007) Borel summability and Lindstedt series. *Commun Math Phys* 269(1):175–193
- Craig W, Wayne CE (1993) Newton's method and periodic solutions of nonlinear wave equations. *Commun Pure Appl Math* 46(11):1409–1498
- Davie AM (1994) The critical function for the semi-standard map. *Nonlinearity* 7(1):219–229
- Eliasson LH (1996) Absolutely convergent series expansions for quasi periodic motions. *Math Phys Electron J* 2(4):1–33
- Gallavotti G (1983) *The elements of mechanics*, Texts and monographs in physics. Springer, New York
- Gallavotti G (1994a) Twistless KAM tori. *Commun Math Phys* 164(1):145–156
- Gallavotti G (1994b) Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review. *Rev Math Phys* 6(3):343–411
- Gallavotti G (2001) Renormalization group in statistical mechanics and mechanics: gauge symmetries and vanishing beta functions, Renormalization group theory in the new millennium III. *Phys Rep* 352(4–6):251–272
- Gallavotti G (2006) Classical mechanics. In: Françoise JP, Naber GL, Tsun TS (eds) *Encyclopedia of mathematical physics*, vol 1. Elsevier, Oxford, UK
- Gallavotti G, Gentile G (2002) Hyperbolic low-dimensional invariant tori and summations of divergent series. *Commun Math Phys* 227(3):421–460
- Gallavotti G, Gentile G, Mastropietro V (1999) A field theory approach to Lindstedt series for hyperbolic tori in three time scales problems. *J Math Phys* 40(12):6430–6472
- Gallavotti G, Bonetto F, Gentile G (2004) *Aspects of ergodic, qualitative and statistical theory of motion*, Texts and monographs in physics. Springer, Berlin
- Gallavotti G, Gentile G, Giuliani A (2006) Fractional Lindstedt series. *J Math Phys* 47(1):1–33
- Gentile G (1995) Whiskered tori with prefixed frequencies and Lyapunov spectrum. *Dyn Stab Syst* 10(3):269–308
- Gentile G (2003) Quasi-periodic solutions for two-level systems. *Commun Math Phys* 242(1–2):221–250
- Gentile G (2006a) Brjuno numbers and dynamical systems. In: *Frontiers in number theory, physics, and geometry*. Springer, Berlin
- Gentile G (2006b) Diagrammatic techniques in perturbation theory. In: Françoise JP, Naber GL, Tsun TS (eds) *Encyclopedia of mathematical physics*, vol 2. Elsevier, Oxford, UK
- Gentile G (2006c) Resummation of perturbation series and reducibility for Bryuno skew-product flows. *J Stat Phys* 125(2):321–361
- Gentile G (2007) Degenerate lower-dimensional tori under the Bryuno condition. *Ergod Theory Dyn Syst* 27(2):427–457
- Gentile G, Gallavotti G (2005) Degenerate elliptic tori. *Commun Math Phys* 257(2):319–362
- Gentile G, Mastropietro V (1996) Methods for the analysis of the Lindstedt series for KAM tori and renormalizability in classical mechanics. A review with some applications. *Rev Math Phys* 8(3):393–444
- Gentile G, Mastropietro V (2001) Renormalization group for one-dimensional fermions. A review on mathematical results. Renormalization group theory in the new millennium III. *Phys Rep* 352(4–6):273–437
- Gentile G, Mastropietro V (2004) Construction of periodic solutions of nonlinear wave equations with Dirichlet boundary conditions by the Lindstedt series method. *J Math Pures Appl* 83(8):1019–1065
- Gentile G, Procesi M (2006) Conservation of resonant periodic solutions for the one-dimensional nonlinear Schrödinger equation. *Commun Math Phys* 262(3):533–553
- Gentile G, van Erp TS (2005) Breakdown of Lindstedt expansion for chaotic maps. *J Math Phys* 46(10):1–20
- Gentile G, Bartuccelli MV, Deane JHB (2005a) Summation of divergent series and Borel summability for strongly dissipative differential equations with periodic or quasiperiodic forcing terms. *J Math Phys* 46(6):1–21
- Gentile G, Cortez DA, Barata JCA (2005b) Stability for quasi-periodically perturbed Hill's equations. *Commun Math Phys* 260(2):403–443
- Gentile G, Mastropietro V, Procesi M (2005c) Periodic solutions for completely resonant nonlinear wave equations with Dirichlet boundary conditions. *Commun Math Phys* 256(2):437–490
- Gentile G, Bartuccelli MV, Deane JHB (2006) Quasiperiodic attractors, Borel summability and the Bryuno condition for strongly dissipative systems. *J Math Phys* 47(7):1–10
- Gentile G, Bartuccelli MV, Deane JHB (2007) Bifurcation curves of subharmonic solutions and Melnikov theory under degeneracies. *Rev Math Phys* 19(3):307–348
- Harary F (1969) *Graph theory*. Addison-Wesley, Reading
- Iserles A, Nørsett SP (1999) On the solution of linear differential equations in Lie groups. *Philos Trans R Soc A Math Phys Eng Sci* 357(1754):983–1019
- Khanin K, Lopes Dias J, Marklof J (2007) Multi-dimensional continued fractions, dynamical renormalization and KAM theory. *Commun Math Phys* 270(1):197–231
- Koch H (1999) A renormalization group for Hamiltonians, with applications to KAM tori. *Ergod Theory Dyn Syst* 19(2):475–521
- Kolmogorov AN (1954) On conservation of conditionally periodic motions for a small change in Hamilton's function (Russian). *Dokl Akad Nauk SSSR* 98:527–530
- Kuksin SB (1993) Nearly integrable infinite-dimensional Hamiltonian systems, *Lecture notes in mathematics*, vol 1556. Springer, Berlin

- Levi-Civita T (1954) *Opere matematiche. Memorie e note*. Zanichelli, Bologna
- MacKay RS (1993) Renormalisation in area-preserving maps, *Advanced series in nonlinear dynamics*, vol 6. World Scientific Publishing, River Edge
- Moser J (1962) On invariant curves of area-preserving mappings of an annulus. *Nachr Akad Wiss Göttingen Math Phys Kl II*:1–20
- Moser J (1967) Convergent series expansions for quasi-periodic motions. *Math Ann* 169:136–176
- Poincaré H (1892–1899) *Les méthodes nouvelles de la mécanique céleste*. Gauthier-Villars, Paris
- Pöschel J (1982) Integrability of Hamiltonian systems on Cantor sets. *Commun Pure Appl Math* 35(5):653–696
- Wintner A (1941) *The analytic foundations of celestial mechanics*. Princeton University Press, Princeton

### Books and Reviews

- Berretti A, Gentile G (2001b) Renormalization group and field theoretic techniques for the analysis of the Lindstedt series. *Regul Chaotic Dyn* 6(4):389–420
- Gentile G (1999) *Diagrammatic techniques in perturbations theory, and applications. Symmetry and perturbation theory*. World Science, River Edge

# Perturbation Theory and the Method of Detuning

Giuseppe Pucacco

Department of Physics, University of Rome Tor Vergata, Rome, Italy

## Article Outline

Glossary

Introduction

Classical Detuning

Normalization

Quasi-resonant Normalization

Detuned Resonant 2-DOF Systems

Variables Adapted to the  $k:\ell$  Resonance

Classical Examples

Future Directions

Bibliography

## Glossary

**Bifurcation** Qualitative change in the phase-space structure of a system associated to the variation of a parameter, e.g., the energy.

**Detuning** Small deviation from exact commensurability of the frequencies of the unperturbed system.

**Normal form** Simplified Hamiltonian for a perturbed near-integrable system constructed by means of a series of canonical transformations.

**Perturbation theory** Set of mathematical tools aimed at obtaining informations on a perturbed non-integrable system starting from the structure of an unperturbed integrable one.

## Introduction

It is well known that most nonlinear dynamical systems cannot be solved by exact analytical

methods and can only be studied through a perturbation approach.

Modern perturbation theory was created by Henry Poincaré and has the transformation to *normal form* as one of its cornerstones. Basically this can be constructively realized by solving a *homological equation* at each level of the perturbation process. The main problem met in solving this equation is the presence of *small denominators*, which are in turn related to *near resonances* in the unperturbed system. In practice, small denominators reduce the domain of analyticity of the normalizing transformation, and they can accumulate so to make this domain vanish. This is not surprising: as remarked by Moser (1973), the normal form (at least for problems of interest in physics and in celestial mechanics) has an additional symmetry and hence conserved quantity, so the transformation bringing the original system into normal form can be analytical only if the original system already had (at least in some neighborhood of the unperturbed solution) such an additional symmetry or conserved quantity. Note also, in this respect, that the most basic criteria for convergence of the normalizing transformation (e.g., the condition that eigenvalues of the unperturbed linear operator lie in a Poincaré domain) are not satisfied in the case of resonant Hamiltonian systems.

It may seem paradoxical that, from this point of view, exactly resonant systems are easier to analyze than near-resonant ones. In fact, while for nonresonant system the normal form is *linear* and thus all nonlinear perturbing terms should be eliminated by solving the homological equation, for resonant system, the normal form contains resonant terms, i.e., we (have to) accept that certain nonlinear terms are not eliminated. This obviously implies that the normal form dynamics is nonlinear and thus at the same time richer and more difficult to analyze; but when we look at the *reduction* to normal form, the fact we accept to keep certain nonlinear terms makes of course things easier since can be exploited to tame the most dangerous small denominators.

This trivial observation is at the basis of the method of *detuning* (Sanders et al. 2007), also going under the name of *frequency deviation* in the Russian literature (Arnol'd 1989). The idea is that, when we have a near-resonance, we consider an exactly resonant unperturbed system and treat the detuning of the frequency as a small deviation, to be included in the perturbing function rather than inscribed in the structure of the unperturbed system. One of the consequences of this trick is that we aim at a *resonant* normal form containing (nonlinear) resonant terms. These are exactly the terms which would call the small denominators into action were we attempting to eliminate them, and it is therefore not surprising that the method can give a much better outcome in terms of convergence of the normalizing transformation (and thus conjugation of the normal form dynamics with that of the full original system).

Aim of this entry is a short review of the detuning method in classical mechanics with particular emphasis in few degrees of freedom near-resonant Hamiltonian systems with exact or approximate discrete symmetries perturbed by homogeneous polynomial terms.

## Classical Detuning

Our main characters are *resonance* and *detuning*: *Resonance* is the ubiquitous phenomenon producing spectacular features of non-integrable Hamiltonian systems; see, e.g., the celebrated Hénon-Heiles (1964) problem; *Detuning* is the simple device that, under certain hypotheses, allows us to explore the phase space of a generic system undergoing resonant behavior (Henrard 1970; Schmidt 1974; Verhulst 1979).

We will limit ourselves to focus on the classical setting with an eye to the semiclassical and quantum applications. Bifurcations and stability/instability transitions are features common to all these frameworks. For systems close to resonance with nonlinear coupling terms, crossing an exact resonance is a generic phenomenon triggered by changes in some parameter, typically either the energy or some control parameter. Let us consider a perturbed Hamiltonian model of the form

$$\mathcal{H}(\mathbf{p}, \mathbf{q}; \varepsilon) = \sum_{j=0}^{\infty} \varepsilon^j \mathcal{H}_j(\mathbf{p}, \mathbf{q}) \quad (1)$$

with  $\mathbf{p}, \mathbf{q}$  canonical coordinates in  $\mathbf{R}^{2N}$  and  $\varepsilon$  a bookkeeping parameter – not necessarily small at this stage – playing the role of ordering the terms in the series. The usual setting is provided by a system around an elliptic equilibrium for which

$$\mathcal{H}_0 = \frac{1}{2} \sum_{a=1}^N \omega_a (p_a^2 + q_a^2) \quad (2)$$

with  $\omega_a \in \mathbf{R}^N$ ;  $a = 1, \dots, N$ . For nondegenerate systems, situations of effective equilibrium can be obtained by exploiting a symmetry through reduction or by expanding around some nominal (usually periodic) solutions.

We say that the frequency vector of the unperturbed system is *resonant* if

$$\sum_{a=1}^N \omega_a n_a = 0, \quad (3)$$

with resonant module provided by the vectors

$$n_a^{(j)} \in \mathbf{Z}^N, \quad a = 1, \dots, N; \quad j = 1, \dots, M. \quad (4)$$

Here  $M$  is the *interaction number* and  $j$  the associated label.

Suppose now that a given pair of components of the frequency vector (without loss of generality we can rearrange indices and take the first two components  $\omega_1, \omega_2$ ) are “close” to a rational ratio  $k/\ell$ ,  $k, \ell \in \mathbf{Z}$ , then we can introduce the *detuning parameter*  $\delta$  by defining

$$\delta := \frac{\omega_1}{\omega_2} - \frac{k}{\ell}. \quad (5)$$

We can assume

$$\delta = O(\varepsilon^r), \quad (6)$$

where the power  $r$  is chosen according to the order of the first resonant term in the normal form. As a matter of principle, with  $N$  degrees of freedom, we can introduce  $N - 1$  detuning parameters.

## Normalization

The most effective way to explore the backbone dynamics of a perturbed Hamiltonian system is to construct a suitable *normal form* (see Broer 2021; Broer and Hanßmann 2021; Giorgilli 2002; Sanders et al. 2007). Among the various definitions of normal form, here we adopt the following one: we find a simplified Hamiltonian series whose Poisson bracket with the unperturbed part (2) is vanishing small in  $\varepsilon^{r+1}$  when  $r$  steps of normalization have been performed. Resonant normal forms are computed when a given *exact* resonance prevents standard normalization algorithms to work. However, the power of the detuning is just to exploit it even in strictly nonresonant cases, when, in principle, a standard Poincaré-Birkhoff normalization would perform flawlessly.

The idea is to proceed as in the case where the unperturbed part is exactly at the interesting resonance and hence leave in the normal form the associated resonant terms. The benefit is twofold: to eliminate several terms possessing *small divisors* and to get an effective tool to explore the phenomena associated with the resonance.

Let us then consider a sequence of polynomial functions in the phase space

$$\{G\} = G_1, G_2, \dots, \quad G_j = O(\varepsilon^j), \quad (7)$$

To each term of the sequence (7) is naturally associated the linear differential operator

$$\exp[\mathcal{L}_{G_j}] = \sum_k \frac{\varepsilon^{jk}}{k!} \mathcal{L}_{G_j^k}, \quad (8)$$

whose action on a generic function  $F$  is given by the Poisson bracket:

$$\mathcal{L}_G F := \{F, G\} = \frac{\partial F}{\partial \mathbf{q}} \cdot \frac{\partial G}{\partial \mathbf{p}} - \frac{\partial F}{\partial \mathbf{p}} \cdot \frac{\partial G}{\partial \mathbf{q}}. \quad (9)$$

Under the action of this operator, the original Hamiltonian system (1) undergoes a canonical transformation to new variables  $(\mathbf{Q}, \mathbf{P})$  such that

$$\mathbf{q} = e^{\mathcal{L}_G} \mathbf{Q}, \quad \mathbf{p} = e^{\mathcal{L}_G} \mathbf{P}; \quad (10)$$

and the new Hamiltonian is

$$\begin{aligned} \mathcal{K}(\mathbf{Q}, \mathbf{P}) &= (e^{\mathcal{L}_G} \mathcal{H})(\mathbf{Q}, \mathbf{P}) \\ &= \mathcal{H}(e^{\mathcal{L}_G} \mathbf{Q}, e^{\mathcal{L}_G} \mathbf{P}), \end{aligned} \quad (11)$$

where every function is assumed to be in the form of power series.

The process is recursive with first step given by  $G_1$  (we can start with the cubic term because a  $G_0$  would give only trivial linear transformations). The general relation (11) takes the form

$$\begin{aligned} \mathcal{K}_0 + \varepsilon \mathcal{K}_1 + \varepsilon^2 \mathcal{K}_2 + \dots \\ = (1 + \varepsilon \mathcal{L}_{G_1} + \dots)(\mathcal{H}_0 + \varepsilon \mathcal{H}_1 + \varepsilon^2 \mathcal{H}_2 + \dots). \end{aligned} \quad (12)$$

By equating polynomials of the same degree in  $\varepsilon$ , we get the system:

$$\begin{aligned} \mathcal{K}_0 &= \mathcal{H}_0, \\ \mathcal{K}_1 &= \mathcal{H}_1 + \mathcal{L}_{G_1} \mathcal{H}_0, \\ \mathcal{K}_2 &= \mathcal{H}_2 + \mathcal{L}_{G_1} \mathcal{H}_1 + \frac{1}{2} \mathcal{L}_{G_1}^2 \mathcal{H}_0, \\ &\vdots = \vdots \end{aligned}$$

and equations involving terms of higher degrees. The first equality simply states that the zero-order new Hamiltonian coincides with the zero-order old (unperturbed) one. The second equation has to be solved to find the first-order term  $\mathcal{K}_1$ : to proceed, we have to make some decision about the structure the new Hamiltonian must have, that is we have to choose a *normal form* for it. We therefore select the new Hamiltonian in such a way that it admits a new integral of motion; that is we take a certain function, say  $F$ , and impose that  $\{\mathcal{K}, F\} = 0$ . As defined above, the usual choice is that of taking  $F = \mathcal{H}_0 = \mathcal{K}_0$  so that (2) plays the double role of determining the specific form of the transformation *and* assuming the status of the second integral of motion.

With this choice, the fundamental equation of the chain, that we can also write in the form

$$\mathcal{K}_1 = \mathcal{H}_1 + \mathcal{L}_{G_1} \mathcal{H}_0 = \mathcal{H}_1 - \mathcal{L}_{\mathcal{H}_0} G_1, \quad (13)$$

is solved by requiring that the action of the operator  $\mathcal{L}_{\mathcal{H}_0}$  on any polynomial function which commutes with  $\mathcal{H}_0$  is to “kill” it, whereas its action on

any other polynomial gives a uniquely defined nonvanishing polynomial. We can therefore split the polynomial  $\mathcal{H}_1$  appearing in (13) as

$$\mathcal{H}_1 = \mathcal{H}_1^K + \mathcal{H}_1^R \quad (14)$$

where  $\mathcal{H}_1^K$  is the part belonging to the *kernel* of  $\mathcal{L}_{\mathcal{H}_0}$ ,

$$\mathcal{L}_{\mathcal{H}_0} \mathcal{H}_1^K \equiv 0, \quad (15)$$

and  $\mathcal{H}_1^R$  is the part belonging to the *range* of  $\mathcal{L}_{\mathcal{H}_0}$ ,

$$\mathcal{L}_{\mathcal{H}_0} \mathcal{H}_1^R = R_1, \quad (16)$$

where  $R_1$  is a nonvanishing cubic polynomial.

Since the new Hamiltonian is in normal form, if and only if it stays in the kernel of  $\mathcal{L}_{H_0}$ , we can then *solve* (13) by applying the simple prescription:

$$\mathcal{K}_1 = \mathcal{H}_1^K, \quad G_1 = \mathcal{L}_{\mathcal{H}_0}^{-1} \mathcal{H}_1^R. \quad (17)$$

We observe that the operation implied in the second expression is well defined and easily solved (Boccaletti and Pucacco 1999; Giorgilli 2002) to find the first Lie generating function  $G_1$ : we can therefore use it in the subsequent equations of the system to compute the terms  $H_j^{(1)}$  with  $j > 1$  (which are still *not* in normal form), and go one step further by expanding Eq. (11) at order 2 and applying the above recipe to compute  $G_2$  and the normal form at order 2

$$\begin{aligned} \mathcal{K}_2 &= \mathcal{H}_2^{(1)} + \mathcal{L}_{G_2} \mathcal{H}_0 \\ &= \mathcal{H}_2 + \mathcal{L}_{G_1} \mathcal{H}_1 + \frac{1}{2} \mathcal{L}_{G_1}^2 \mathcal{H}_0 + \mathcal{L}_{G_2} \mathcal{H}_0. \end{aligned} \quad (18)$$

The procedure can then be iterated up to an arbitrary order. It is usually stopped before a threshold order  $N_0$  at which the *remainder*

$$\left| \sum_{j=N_0+1}^{\infty} \mathcal{H}_j^{(N_0)} \right|,$$

ceases to decrease.

## Quasi-resonant Normalization

We have so far discussed the general normalization scheme; in this work, we are specially interested in the *quasi-resonant case*, and in this case, the normalization algorithm may present some technical issue.

- The most natural way to construct a detuned resonant normal form is simply to exploit the standard method of recursive solution of the homological equation in which the linear operator is associated to the nonresonant linear part, but the choice of the resonant terms left in the normal form is dictated by the quasi-resonance of interest (Method 1).
- In alternative, terms proportional to the detuning (with a given bookkeeping order) can be treated as perturbation terms and treated, at each step, by the resonant algorithm with exactly resonant homological operator (Method 2).

Each of these methods has pros and cons:

- Method 1 is straightforward since it does not necessitate any adaptation but, in general, produces many terms with progressively smaller denominators. Moreover, it is computationally more demanding since the memory occupancy is larger than in the other approach. This can be a serious problem when the order of normalization is quite high.
- Method 2 requires to be adapted to manage terms of different *algebraic order* corresponding to the same *perturbation order*. By a proper adjustment of bookkeeping orders in the algorithm, the treatment of terms of different degree at each step can be efficiently coded and so this method proves to be quite fast and accurate.

We remark that, in order to compare the results obtained in the two approaches, the simplest way is to perform, in the outcome of Method 1, a series expansion in the detuning and a rearrangement of terms by means of powers of the bookkeeping parameter.

## Detuned Resonant 2-DOF Systems

The reference case, which can be considered with no loss of generality, is provided by a two degrees-of-freedom system around an elliptic equilibrium like (2) with perturbation given by a series of homogeneous polynomials. It is convenient to pass to action-angle variables  $(\mathbf{J}, \boldsymbol{\phi})$ ; with these, the Hamiltonian can be written in the form of the perturbed oscillator

$$\mathcal{H} = \omega_1 J_1 + \omega_2 J_2 + \sum_{j=1}^{\infty} \varepsilon^j \mathcal{H}_j(\mathbf{J}, \boldsymbol{\phi}). \quad (19)$$

In the case where

$$\frac{\omega_1}{\omega_2} \notin \mathbf{Q},$$

only the solutions  $J_1 = 0$  and  $J_2 = 0$  are periodic. If

$$\frac{\omega_1}{\omega_2} \in \mathbf{Q},$$

then *all* orbits are periodic. Solutions for  $J_1 = 0$  and for  $J_2 = 0$  are axial oscillations: they are called *normal modes*. In phase space, the energy manifold is an ellipsoid diffeomorphic to the sphere  $\mathbf{S}^3$ :

$$\omega_1 J_1 + \omega_2 J_2 = E_0.$$

When the oscillator is perturbed, the orbits will have different evolutions: let us still call normal modes the solutions, if they exist, corresponding to  $J_1 = 0$  and  $J_2 = 0$  (these are *nonlinear normal modes* in the sense of Moser (1970) and Weinstein (1973)) and *orbits in generic position* those for which both  $J_1$  and  $J_2$  are nonzero).

Consider the detuned  $k:\ell$  resonance defined by (5). It will be convenient to use complex coordinates

$$x_k = q_k + i p_k, \quad y_k = q_k - i p_k; \quad (20)$$

given a term

$$C_n x_1^{n_1} y_1^{n_2} x_2^{n_3} y_2^{n_4}, \quad \sum_{j=1}^4 n_j = N$$

of a polynomial  $\mathcal{H}_N$  of degree  $N$ , the solutions of eqs. (17) are provided by the vanishing of the scalar product

$$k(n_1 - n_2) + \ell(n_3 - n_4). \quad (21)$$

If (21) vanishes, the corresponding term belongs to the normal form; if not, it produces a term in the generating function  $G_{N-2}$ . The resonant normal form therefore is (Sanders et al. 2007)

$$\begin{aligned} \mathcal{K} = & \frac{1}{2}(\omega_1 x_1 y_1 + \omega_2 x_2 y_2) \\ & + \varepsilon^2 [\alpha_1 x_1^2 y_1^2 + \alpha_2 x_2^2 y_2^2 + \alpha_3 x_1 y_1 x_2 y_2] \\ & + \varepsilon^{k+\ell-2} 2\alpha_4 [x_1^\ell y_2^k + x_2^k y_1^\ell] + \dots \end{aligned} \quad (22)$$

with  $k, \ell \in \mathbf{N}$ ,  $\alpha_j \in \mathbf{R}$ ,  $j = 1, \dots, 4$ . In action-angle variables

$$\begin{aligned} x_a = & -i\sqrt{2J_a}e^{i\phi_a}, \quad y_a = \sqrt{2J_a}e^{-i\phi_a}, \quad a \\ = & 1, 2, \end{aligned} \quad (23)$$

we have

$$\begin{aligned} \mathcal{K} = & \omega_1 J_1 + \omega_2 J_2 + \varepsilon^2 (\alpha_1 J_1^2 + \alpha_2 J_2^2 + \alpha_3 J_1 J_2) + \dots \\ & + 2^{(k+\ell+2)/2} \varepsilon^{k+\ell-2} \alpha_4 J_1^{\ell/2} J_2^{k/2} \cos(\ell\phi_1 - k\phi_2). \end{aligned} \quad (24)$$

## Variables Adapted to the $k:\ell$ Resonance

Let  $(k^*, \ell^*)$  such that  $(k, \ell) \cdot (k^*, \ell^*) = v \neq 0$ , with  $v \in \mathbf{Z}$ . We will now construct the *variables adapted to the resonance  $k/\ell$* . Define the matrix

$$\widehat{M} = \begin{pmatrix} \ell & -k \\ k^* & \ell^* \end{pmatrix}$$

and put



$$\mathcal{E} = kJ_1 + \ell J_2.$$

The coordinates adapted to the  $k/\ell$ -resonance are  $\{r_1, r_2, \psi_1, \psi_2\}$  and are defined by

$$\begin{aligned}\psi_1 &= \ell\phi_1 - k\phi_2, & \psi_2 &= k^*\phi_1 + \ell^*\phi_2, \\ J_1 &= \frac{1}{v}(\ell r_1 + k^*r_2) = \frac{1}{v}(\ell \mathcal{R} + k^*\mathcal{E}), \\ J_2 &= \frac{1}{v}(-k r_1 + \ell^*r_2) = \frac{1}{v}(-k \mathcal{R} + \ell^*\mathcal{E}).\end{aligned}\quad (25)$$

In these coordinates

$$\begin{aligned}K &= (\omega_1 k^* + \omega_2 \ell^*)\mathcal{E} + \delta R \\ &+ \varepsilon^{k+\ell-2} 2^{(k+\ell+2)/2} v^{-(k+\ell)/2} \alpha_4 (\ell \mathcal{R} + k^*\mathcal{E})^{\frac{\ell}{2}} (-k \mathcal{R} + \ell^*\mathcal{E})^{\frac{k}{2}} \cos \psi_1 \\ &+ \varepsilon^2 [(\ell \Delta_1 - k \Delta_2) \mathcal{R}^2 + 2(k^* \Delta_1 + \ell^* \Delta_2) \mathcal{R} \mathcal{E} + (k^* \Delta_1^* + \ell^* \Delta_2^*) \mathcal{E}^2] \\ &+ \dots\end{aligned}\quad (26)$$

with

$$\begin{aligned}\Delta_1 &= \begin{vmatrix} \alpha_1 & k \\ \alpha_3/2 & \ell \end{vmatrix} & \Delta_2 &= \begin{vmatrix} \alpha_3/2 & k \\ \alpha_2 & \ell \end{vmatrix} \\ \Delta_1^* &= \begin{vmatrix} \alpha_1 & k^* \\ \alpha_3/2 & -\ell^* \end{vmatrix} & \Delta_2^* &= \begin{vmatrix} \alpha_3/2 & k^* \\ \alpha_2 & -\ell^* \end{vmatrix}.\end{aligned}$$

The detuning term appears explicitly and the angle  $\psi_2$  is not present in (26), consistent with  $\mathcal{E}$  being a conserved quantity for the resonant normal form.

The general analysis of the global structure of phase space of each resonant normal form proceeds with the study of the bifurcations induced by changes in the internal “distinguished” parameter ( $\mathcal{E}$ ) and in the external or “control” parameters ( $\delta, \alpha_j$ ). It can be approached by following two methods: (a) the analytical approach based on singularity theory considering (26) as a universal unfolding of a “catastrophe” germ (Golubitsky and Schaeffer 1985, Arnold 1994, Marchesiello and Pucacco 2014); (b) the geometric approach using the Poisson algebra of invariants of the harmonic oscillator and the corresponding reduced systems (Cushman and Bates 1997, Pucacco and Marchesiello 2014). Clear and accurate references for these methods are given by Efstathiou (2005) and Hanßmann (2007).

## Classical Examples

In this section, we relate the bifurcations of the  $k/\ell$ -resonances to their universal unfolding and

catastrophe maps. Rather than providing a general formal treatment, we find more convenient to explicitly work out the most representative cases in 2 DOF.

### The Symmetric 1:1 Resonance

The expression (22) cannot be directly applied to the  $k = \ell = 1$  case because additional quadratic terms with many other free parameters should be included. However, when enforcing typical symmetries of physical models, the normal form is simplified and belongs to the family (24). The standard case is that of the *reflection symmetries* with respect to one or both normal modes (Tuwankotta and Verhulst 2000). In this case, it is simpler to assume  $k = \ell = 2$  so that the normal form is

$$\begin{aligned}\mathcal{K} &= J_1 + J_2 + \varepsilon^2 (\delta J_1 + \alpha_1 J_1^2 + \alpha_2 J_2^2 + \alpha_3 J_1 J_2) \\ &+ 2\varepsilon^2 \alpha_4 J_1 J_2 \cos(2\phi_1 - 2\phi_2) + \dots,\end{aligned}\quad (27)$$

where an obvious rescaling has been applied but, for simplicity, we have kept the same symbols for the rescaled control parameters. In terms of variables adapted to the resonance we have, with the choice

$$\hat{M} = \begin{pmatrix} 2 & -2 \\ 2 & 2 \end{pmatrix},\quad (28)$$

and setting now  $\varepsilon = 1$ , the function

$$\begin{aligned} \mathcal{K} = & (1 + \Delta)\mathcal{E} + A_+\mathcal{E}^2 + (B\mathcal{E} + \Delta)\mathcal{R} \\ & + A_-\mathcal{R}^2 + C\sqrt{\mathcal{E}^2 - \mathcal{R}^2}\cos 2\psi, \end{aligned} \quad (29)$$

where

$$\begin{aligned} A_{\pm} &:= \frac{\alpha_1 + \alpha_2 \pm \alpha_3}{4}, \quad B := \frac{\alpha_1 - \alpha_2}{2}, \\ C &:= \frac{\alpha_4}{2}, \\ \Delta &:= \frac{\delta}{2}. \end{aligned} \quad (30)$$

Singularity theory is implemented by finding an unfolding of the central singularity with double reflection symmetry in the plane. By introducing coordinates in the plane defined by

$$x = \sqrt{\mathcal{E} + \mathcal{R}} \cos \psi, \quad y = \sqrt{\mathcal{E} + \mathcal{R}} \sin \psi, \quad (31)$$

so that

$$\mathcal{R} = x^2 + y^2 - \mathcal{E}, \quad (32)$$

we get the function (Broer et al. 1998; Marchesello and Pucacco 2014)

$$\begin{aligned} F(x, y; \varepsilon_{\pm}, \mu, u_{\pm}) = & \varepsilon_-x^4 + \mu x^2y^2 + \varepsilon_+y^4 \\ & + u_-x^2 + u_+y^2 \end{aligned} \quad (33)$$

where

$$\begin{aligned} \varepsilon_{\pm} &:= A \pm C, \quad u_{\pm} := \Delta - (2A - B \pm 2C)\mathcal{E}, \\ \mu &:= 2A, \quad A := A_-. \end{aligned} \quad (34)$$

This is the standard form of the unfolding of the *cusp* catastrophe (Arnold 1994) giving pitchfork bifurcations. The bifurcation set is determined by finding the critical points of  $F$  inside the “limit circle”

$$x^2 + y^2 = 2\mathcal{E}.$$

We find four conditions corresponding to pairs of critical points colliding with either the origin or the limit circle:

$$\begin{aligned} u_+ &= 0, \\ u_- &= 0, \\ Au_- - (A - C)u_+ &= 0, \\ (A + C)u_- - Au_+ &= 0. \end{aligned}$$

Taking into account (34), we see that a nonvanishing detuning produces bifurcation thresholds for the distinguished parameter  $\mathcal{E}$  (the “energy”).

More precisely, at

$$\mathcal{E}_{1,2}^L := \frac{\Delta}{\pm 2(A + C) - B} \quad (35)$$

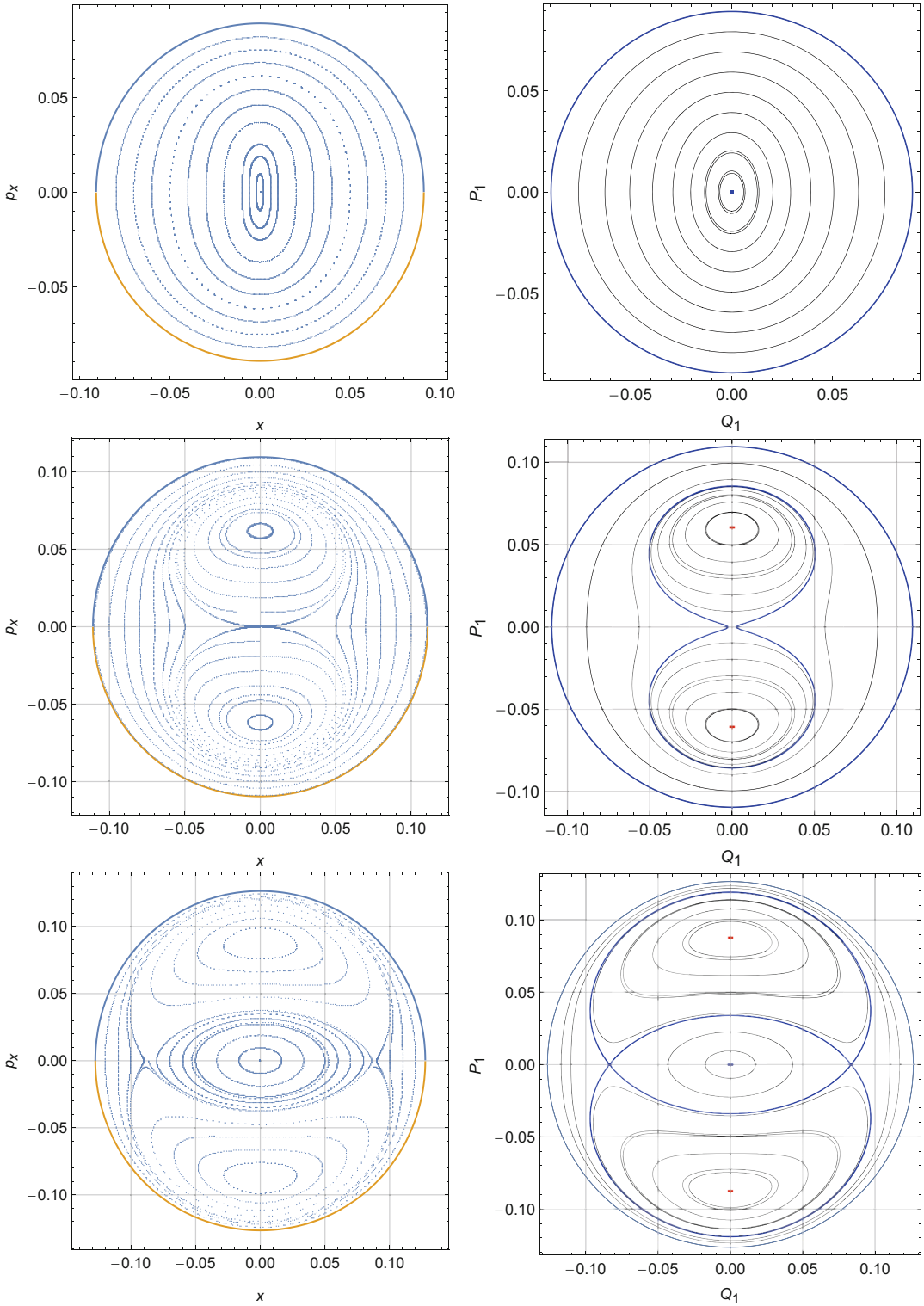
the “loop” families ( $\psi = \pm \pi/2$ ) bifurcate respectively from the  $J_1, J_2$  normal modes, and at

$$\mathcal{E}_{1,2}^I := \frac{\Delta}{\pm 2(A - C) - B} \quad (36)$$

the “inclined” families ( $\psi = 0, \pi$ ) bifurcate respectively from the  $J_1, J_2$  normal modes. In Fig. 1, we see an explicitly worked example, useful to appreciate the power of the detuning approach: we examine a quasi-isotropic oscillator with  $\delta = -0.0125$  perturbed by quartic terms. The bifurcation of the first inclined family occurs at  $\mathcal{E}_1^I$  and that of the first loop (unstable in the present instance) at  $\mathcal{E}_1^L$ . The detuned resonant normal form given by formula (27) (right panels) is able to reproduce quite accurately the phase-space structure of the physical system (left panels) in a wide range of  $\mathcal{E}$ . A nonresonant normal form would have displayed only the torus foliation of the upper right panel. A standard resonant normal form would have captured only either the first or the second bifurcation.

A complementary view to the algebraic setting above is offered by the geometric approach based on the invariants of the isotropic oscillator. It can be proven (Cushman and Bates 1997) that the Hilbert basis of the functions in  $\mathbb{R}^4$  invariant under the action generated by the dynamics of the isotropic oscillator is given by

$$\begin{aligned} I_0 &= i(x_1y_1 + x_2y_2) = J_1 + J_2 = \mathcal{E}, \\ I_1 &= i(x_1y_2 + x_2y_1) = 2\sqrt{J_1J_2} \cos \psi, \\ I_2 &= x_1y_2 - x_2y_1 = 2\sqrt{J_1J_2} \sin \psi, \\ I_3 &= i(x_1y_1 - x_2y_2) = J_1 - J_2. \end{aligned}$$



**Perturbation Theory and the Method of Detuning,**  
**Fig. 1** Phase-space structure of a perturbed oscillator with  $\delta = -0.0125$  and a quartic perturbation. The left panels

display numerical Poincaré sections; the right panels display the corresponding analytical sections computed with the normal form (27)

There is a relation among the invariants: the “syzygy” (Hilbert 1993)

$$I_1^2 + I_2^2 + I_3^2 = I_0^2, \quad (37)$$

which can be interpreted as defining the *reduced phase space*.

In terms of the invariants, the Hamiltonian is (Pucacco and Marchesiello 2014)

$$\begin{aligned} h &:= \mathcal{K} - (1 + \Delta)\mathcal{E} - A_+\mathcal{E}^2 \\ &= (B\mathcal{E} + \Delta)I_3 + AI_3^2 + C(I_1^2 - I_2^2). \end{aligned} \quad (38)$$

The reduced dynamics is now determined by the intersection of the two surfaces (37) and (38). Critical points of the reduced dynamics (periodic orbits of the original system) are given by the tangency conditions: they coincide with (35) and (36). In the space of control parameter, the global bifurcation picture is obtained by introducing the combinations (Rose and Kellman 1996, Svitak et al. 2002)

$$\Sigma := -\frac{B\mathcal{E} + \Delta}{2A\mathcal{E}}, \quad \Gamma := \frac{C}{A}. \quad (39)$$

In terms of these, the bifurcation thresholds (35) and (36) become

$$\Sigma(\mathcal{E}_{1,2}^L) = \pm \left(1 + \frac{C}{A}\right), \quad \Sigma(\mathcal{E}_{1,2}^I) = \pm \left(1 - \frac{C}{A}\right). \quad (40)$$

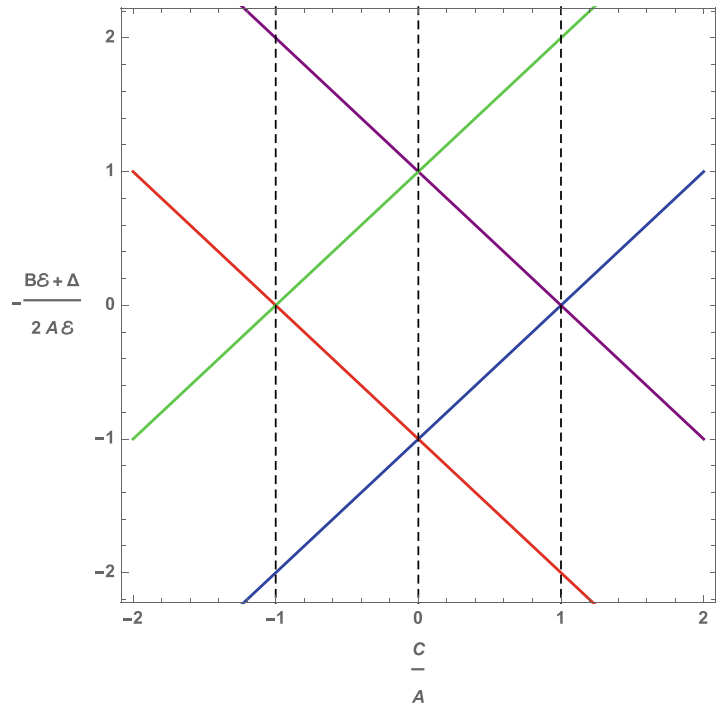
They are plotted in Fig. 2 (taken from Pucacco and Marchesiello 2014), also known as “catastrophe map.” Figure 3 displays the *energy-momentum-map* (Cushman et al. 2007, Schmidt and Dullin 2010) obtained by drawing, in the  $(\mathcal{E}, h)$ -plane, the corresponding bifurcation curves in the case  $|\Gamma| > 1$ .

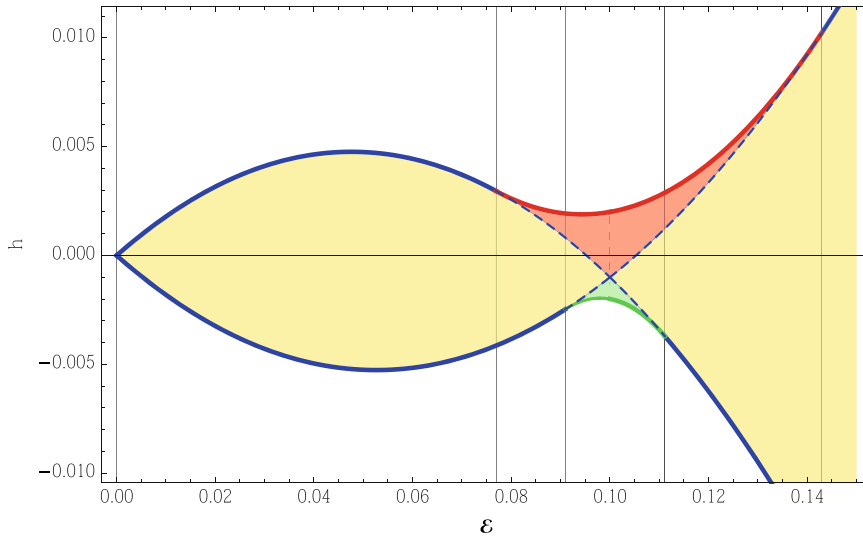
### The 2:1 Resonance

Even more basic than the synchronous resonance seen above, the 2:1 resonance plays an ubiquitous role in many areas spanning from galactic dynamics (Contopoulos 2004, Hanßmann et al. 2020) to molecular physics (Joyeux 1998), where is known as “Fermi resonance.” In classical mechanics, the prototypical system is the spring pendulum (Broer et al. 1998).

### Perturbation Theory and the Method of Detuning,

**Fig. 2** Global representation of the bifurcation sequences of the 1:1 resonance: vertical straight lines are covered by varying  $\mathcal{E}$  and cross the bifurcation lines associated with  $\mathcal{E}_{1,2}^L$  (Eq. (35), green and red lines) and  $\mathcal{E}_{1,2}^I$  (Eq. (36), blue and purple lines). Any point is structurally stable except the degenerate four corners of the inner square (Pucacco and Marchesiello 2014)





**Perturbation Theory and the Method of Detuning, Fig. 3** Image of the energy-momentum map, obtained by drawing, in the  $(\mathcal{E}, h)$ -plane, the corresponding bifurcation

curves, of the symmetric 1:1 resonance in the case  $|\Gamma| = |C/A| > 1$

It can be described with a very simple first-order normal form; however, its explicit treatment is slightly more involved algebraically. The normal form is

$$\begin{aligned} \mathcal{K} = & (2 + \delta)J_1 + J_2 \\ & + 2\epsilon \alpha \sqrt{2J_1} J_2 \cos(\phi_1 - 2\phi_2) \\ & + \dots \end{aligned} \quad (41)$$

The first normal mode becomes unstable at an energy level which depends quadratically on the detuning (see (46) below), when a “banana-shaped” resonant orbit appears in a period-doubling bifurcation. The second normal mode is always stable at first order. By using stereographic variables (Kummer 1976)

$$\begin{cases} x = -\sqrt{2J_1} \cos(\phi_1 - 2\phi_2), \\ y = -\sqrt{2J_1} \sin(\phi_1 - 2\phi_2), \end{cases} \quad (42)$$

so that, as before, the “critical circle”

$$x^2 + y^2 = 2\mathcal{E}$$

delimits the dynamics, we get the function

$$\begin{aligned} F(x, y; \delta, \mathcal{E}, \alpha) = & \delta(x^2 + y^2) \\ & - 2\alpha[2\mathcal{E} - (x^2 + y^2)]x. \end{aligned} \quad (43)$$

This is the standard form of the unfolding of the *fold* catastrophe (Golubitsky and Schaeffer 1985, Arnold 1994). The transcritical bifurcation is now determined by finding the critical points of  $F$  inside the limit circle. We find four solutions:

$$\begin{aligned} x_{1,2} = & \frac{1}{6\alpha} \left( -\delta \pm \sqrt{24\alpha^2\mathcal{E} + \delta^2} \right), \quad y_{1,2} \\ = & 0, \end{aligned} \quad (44)$$

$$x_{3,4} = -\frac{\delta}{2\alpha}, \quad y_{3,4} = \pm \frac{1}{2\alpha} \sqrt{8\alpha^2\mathcal{E} - \delta^2}. \quad (45)$$

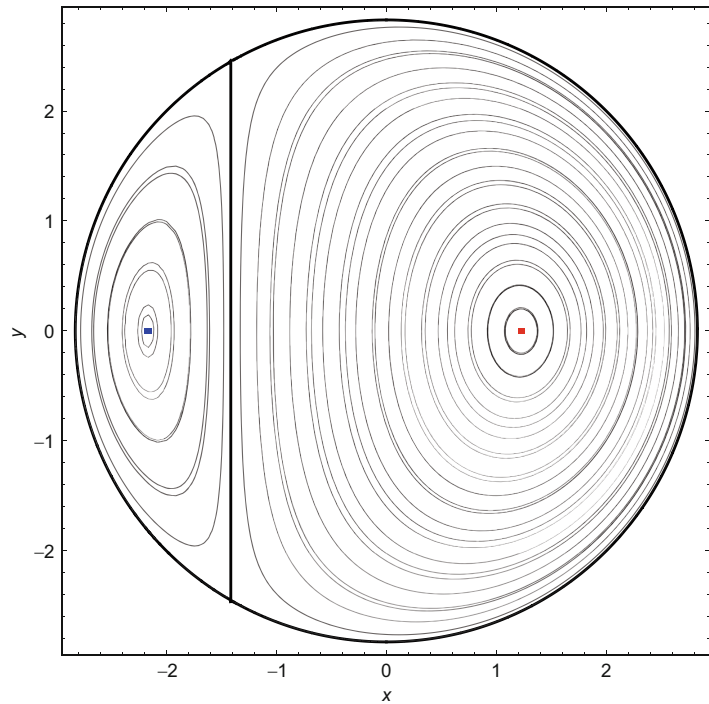
The point  $(x_1, y_1)$  represents the second normal mode and  $(x_2, y_2)$  gives the newly bifurcated orbit. It enters the limit circle when

$$\mathcal{E} > \mathcal{E}_c = \frac{\delta^2}{8\alpha^2}. \quad (46)$$

At this critical energy, the two symmetric points  $(x_3, y_3)$  and  $(x_4, y_4)$  take real values and correspond

## Perturbation Theory and the Method of Detuning,

**Fig. 4** Poincaré section for the 2:1 resonance after the bifurcation of the first normal mode (thick line): the new periodic orbit in generic position is the blue dot, and the second normal mode is the red dot



to the intersection of the limit circle with its stable/unstable manifold; this is the vertical segment of Fig. 4, which can also be interpreted as the Poincaré section in the  $(q_1, p_1)$ -phase plane.

specially suit to analyze *crossing a resonance*. On these basis, it is natural to envision an ever more systematic use of this method in the study of applied nonlinear dynamical systems (Celletti 2021, Gaeta 2021, Panati 2021).

## Future Directions

The detuning approach captures a very relevant qualitative feature: while the dynamics of non-resonant systems (and hence their small enough perturbations) is described by invariant tori of maximal dimension, in resonant system, we will have invariant tori of lower dimension, possibly closed orbits, and these may – suitably deformed – survive perturbations (Farantos et al. 2009, Sadovskii and Zhilinskii 2007). Another important feature emerges when one has to deal with (slowly) *varying parameters*, in particular varying frequencies for the unperturbed system: through the detuning approach, one can consider intervals of parameters and in particular the case in which these go through a resonance (Neishtadt 1990, 2014). In other words, the detuning approach is

## Bibliography

### Primary Literature

- Arnold VI (1989) Mathematical methods of classical mechanics. Springer, Berlin
- Arnold VI (1994) Dynamical systems V, bifurcation theory and catastrophe theory (E.M.S. vol 5). Springer, Berlin
- Boccaletti D, Pucacco G (1999) Theory of orbits. Springer, Berlin
- Broer HW (2021) Normal forms in perturbation theory. ECSS, this volume, Springer
- Broer HW, Hanßmann H (2021) Hamiltonian perturbation theory (and transition to chaos), ECSS, this volume, Springer
- Broer HW, Lunter GA, Vegter G (1998) Equivariant singularity theory with distinguished parameters: two case studies of resonant Hamiltonian systems. *Physica D* 112:64–80
- Celletti A (2021) Perturbation theory in celestial mechanics, ECSS, this volume, Springer

- Contopoulos G (2004) Order and chaos in dynamical astronomy. Springer, Berlin, Heidelberg
- Cushman RH, Bates LM (1997) Global aspects of classical integrable systems. Birkhäuser Basel, Birkhäuser
- Cushman RH, Dullin HR, Hanßmann H, Schmidt S (2007) The  $1:\pm 2$  resonance. *Regular Chaotic Dyn* 12:642–663
- Efstathiou K (2005) Metamorphoses of Hamiltonian systems with symmetries, *Lecture Notes in Mathematics*, 1864. Springer, Berlin
- Farantos SC, Schinke R, Guo H, Joyeux M (2009) Energy localization in molecules, bifurcation phenomena, and their spectroscopic signatures: the global view. *Chem Rev* 109:4248–4271
- Gaeta G (2021) Introduction to perturbation theory, ECSS, this volume, Springer
- Giorgilli A (2002) Notes on exponential stability of Hamiltonian systems. *Centro di Ricerca Matematica E. De Giorgi*, Pisa
- Golubitsky M, Schaeffer DG (1985) Singularities and groups in bifurcation theory, 1, *Applied mathematical sciences*, 51, Springer, Berlin
- Hanßmann H (2007) Local and semi-local bifurcations in Hamiltonian dynamical systems. Springer, Berlin
- Hanßmann H, Marchesiello A, Pucacco G (2020) On the detuned 2:4 resonance. *J Nonlinear Sci* 20:2513–2544
- Hénon M, Heiles C (1964) The applicability of the third integral of motion: some numerical experiments. *Astron J* 69:73–79
- Henrard J (1970) Periodic orbits emanating from a resonant equilibrium. *Celest Mech* 1:437–466
- Hilbert D (1993) *Theory of algebraic invariants*. Cambridge University Press, Cambridge
- Joyeux M (1998) Gustavson's procedure and the dynamics of highly excited vibrational states. *J Chem Phys* 109: 2111–2122
- Kummer M (1976) On resonant non linearly coupled oscillators with two equal frequencies. *Commun Math Phys* 48:53–79
- Marchesiello A, Pucacco G (2014) Equivariant singularity analysis of the 2:2 resonance. *Nonlinearity* 27:43–66
- Moser J (1970) Regularization of Kepler's problem and the averaging method on a manifold. *CommPure Appl Math* 23:609–636
- Moser J (1973) Stable and random motions in dynamical systems. Princeton University Press, Princeton
- Neishtadt AI (1990) Averaging and passage through resonance. *International Congress of Mathematicians*, Kyoto
- Neishtadt AI (2014) Averaging, passage through resonances, and capture into resonance in two-frequency systems. *Russ Math Surv* 69:771
- Panati G (2021) Perturbation theory and molecular dynamics, ECSS, this volume, Springer
- Pucacco G, Marchesiello A (2014) An energy-momentum map for the time-reversal symmetric 1:1 resonance with  $Z_2 \times Z_2$  symmetry. *Physica D* 271:10–18
- Rose JP, Kellman ME (1996) Bending dynamics from acetylene spectra: normal, local, and precessional modes. *J Chem Phys* 105:10743–10754
- Sadovskii DA, Zhilinskii BI (2007) Hamiltonian systems with detuned 1:1:2 resonance. *Manifestations of bidromy*. *Ann Phys* 322:164–200
- Sanders JA, Verhulst F, Murdock J (2007) *Averaging methods in nonlinear dynamical systems*. Springer, Berlin
- Schmidt DS (1974) Periodic solutions near a resonant equilibrium of a Hamiltonian system. *Celest Mech* 9: 81–103
- Schmidt S, Dullin HR (2010) Dynamics near the  $p : -q$  resonance. *Physica D* 239:1884–1891
- Svitak JF, Tyng V, Kellman ME (2002) Bifurcation analysis of higher  $m:n$  resonance spectroscopic Hamiltonian. *J Phys Chem A* 106:10797–10805
- Tuwankotta JM, Verhulst F (2000) Symmetry and resonance in Hamiltonian systems, *SIAM. J Appl Math* 61: 1369–1385
- Verhulst F (1979) Discrete symmetric dynamical systems at the main resonances with applications to axisymmetric galaxies. *R Soc Philos Trans A (London)* 290:435–465
- Weinstein A (1973) Normal modes for nonlinear Hamiltonian systems. *Invent Math* 20:47–58

## Books and Reviews

- Arnold VI, Kozlov VV, Neishtadt AI (1988) *Dynamical systems III: classical mechanics*, (E.M.S. vol.3), Springer, Berlin
- Cicogna G, Gaeta G (1999) *Symmetry and perturbation theory in nonlinear dynamics*. Springer, Berlin
- Landau LD, Lifshitz EM (1960) *Mechanics*. Pergamon Oxford, Pergamon
- Meyer KR, Hall GR, Offin D (2009) *Introduction to Hamiltonian dynamical systems and the N-body problem*. Springer, Berlin



# Computational Methods in Perturbation Theory

Àngel Jorba

Departament de Matemàtiques i Informàtica,  
Universitat de Barcelona, Barcelona, Spain

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
The Solar System  
Dynamics Near an Equilibrium Point of a Hamiltonian System  
Time-Dependent Perturbations  
Quasi-Periodic Motions and KAM Theory  
The Parametrization Method  
Other Situations  
Future Directions  
Bibliography

## Glossary

**Action-angle variables** Conjugate set of coordinates for which an integrable Hamiltonian depends only on the actions and, hence, becomes readily integrable.

**Canonical transformation** A canonical transformation is a change of variables that preserve the Hamiltonian form, that is, it is equivalent to perform the change of variables on the Hamiltonian or on the differential equations.

**Conformally symplectic systems** Conformally symplectic systems model dissipative mechanical systems with a friction which is proportional to the velocity. In contrast to conservative systems, dissipative systems have attractors.

**Delay differential equation** A differential equation in which the derivative of the unknown

function depends not only of the actual state but also of previous states.

**Effective stability** A system is considered effectively stable if the time needed to observe significant changes is longer than the expected lifetime of the system itself. The study of the effective stability is part of the Nekhoroshev Theory.

**First integral** Function that stays constant on each orbit of a differential equation. If there exists a sufficient number of them, the problem becomes integrable.

**Hamiltonian system** This is a specific class of dynamical systems that are governed by differential equations that can be written as  $\dot{q} = \frac{\partial H}{\partial p}$ ,  $\dot{p} = -\frac{\partial H}{\partial q}$ , where  $q = (q_1, \dots, q_\ell)$  are the positions,  $p = (p_1, \dots, p_\ell)$  are the momenta,  $\ell$  is the number of degrees of freedom, and  $H = H(q, p)$  is a smooth function called the Hamilton function or Hamiltonian.

**Invariant manifold** Smooth manifold that is invariant by the dynamics. Jointly with fixed points and periodic orbits (and, sometimes, quasi-periodic orbits) they organize many aspects of the dynamics of the system.

**KAM Theory** This is a collection of results about the preservation of quasi-periodic trajectories under perturbation. The name come from the initials of A.N. Kolmogorov, V.I. Arnol'd, and J. Moser who were the pioneers in this field.

**Nekhoroshev Theory** This is a collection of stability results for an open set of initial conditions and for a finite but very long time span, usually exponentially long with respect to the perturbation.

**Normal form** A simplified version of the equations of motion obtained by means of changes of coordinates.

**Quasi-periodic function** A function  $f$  depending on a variable  $t \in \mathbb{R}$  is said to be quasi-periodic with  $r$  basic frequencies if there exists a function  $F$  defined on  $\mathbb{T}^r$  and a vector  $\omega \in \mathbb{R}^r$  such that  $f(t) = F(\omega t)$ . The components of  $\omega$  are supposed to be linearly independent over the

rational, otherwise the function  $F$  can be rewritten with a lower value for  $r$ . Note that, if  $F$  is smooth and injective, its image is diffeomorphic to a torus of dimension  $r$ .

**Restricted Three-Body Problem** The problem consists in describing the motion of an infinitesimal particle moving under the attraction of two massive bodies. As the particle does not attract the masses, these follow a Keplerian orbit, usually circular (Circular RTBP or CRTBP) or elliptical (Elliptic RTBP or ERTBP).

## Definition of the Subject

The equations of motion of many physical systems can be written as the sum of two parts, a first part that is solvable by closed formulas and a second one which is small. Perturbation theory encloses a set of methods to find approximate solutions that are written as a solution of the solvable part plus small corrections coming from the small part. Numerical methods in perturbation theory comprise numerical techniques that compute approximate solutions by taking advantage of the perturbative situation.

## Introduction

Nature is full of situations that can be approximated by simple models that can be solved “by hand,” using closed expressions. This fact has been extremely useful to understand the basic properties of many physical systems. Better models can be constructed by adding smaller contributions (perturbations) and looking at the effect that these perturbations have on the simple (unperturbed) system. A classical example is given by the motion of the planets in the Solar system. The simple model is obtained by skipping the gravitational attraction between planets which reduces the dynamics to that of several two-body problems (Sun and planet), which can be solved and give rise to the Kepler laws. The perturbation is given by the interactions between planets.

Perturbation theory is a set of methods and techniques to find approximate solutions by starting from solutions of the unperturbed

problem and correcting them to obtain better approximations of the full problem. Usually, the approximated solution as a formal power expansion of a small parameter, where the small parameter represents the size of the perturbing term. This chapter surveys numerical techniques that can be applied in combination with perturbation methods to obtain, in an effective way, approximate solutions of the original problem.

As a motivating example, let us consider the following differential equation,

$$\begin{aligned}\dot{x} &= (A + \varepsilon Q(\theta))x, \\ \dot{\theta} &= \omega,\end{aligned}\tag{1}$$

where  $A$  is a constant  $d \times d$  matrix,  $Q$  is a  $d \times d$  matrix taking values on  $\mathbb{T}^r$ ,  $\omega \in \mathbb{R}^r$  and  $\varepsilon$  is a small but fixed value. The goal is to “solve” this equation, that is, to find a change of variables  $x = P(\theta)y$  that reduces this linear system to constant coefficients,

$$\begin{aligned}\dot{y} &= B(\varepsilon)y, \\ \dot{\theta} &= \omega,\end{aligned}\tag{2}$$

where  $B$  does not depend on  $\theta$ . This allows to know, for instance, the stability of the origin of (1) or to write explicitly a fundamental matrix of (1) as  $P(\omega t) \exp(Bt)$ . To start, let us denote by  $\overline{Q}$  the average of  $Q$ ,

$$\overline{Q} = \frac{1}{(2\pi)^r} \int_{\mathbb{T}^r} Q(\theta) d\theta,$$

and let us define  $\tilde{Q}$  as  $Q - \overline{Q}$ , and  $A_0(\varepsilon) = A + \varepsilon \overline{Q}$ . Then let us write (1) as

$$\begin{aligned}\dot{x} &= (A_0(\varepsilon) + \varepsilon \tilde{Q}(\theta))x, \\ \dot{\theta} &= \omega.\end{aligned}\tag{3}$$

Assume it is possible to compute a  $d \times d$  matrix  $P_1(\omega t)$  such that

$$\dot{P}_1 = A_0 P_1 - P_1 A_0 + \tilde{Q}(\omega t).$$

A sufficient condition for  $P_1$  to exist is that the eigenvalues of  $A_0$ ,  $\lambda_1, \dots, \lambda_d$ , satisfy a suitable non-resonance condition,

$$\left| \lambda_i - \lambda_j + \langle k, \omega \rangle \sqrt{-1} \right| \geq \frac{c}{|k|^\gamma}, \quad \forall k \in \mathbb{Z}^r \setminus \{0\},$$

and for all  $i \neq j$ .

for some  $c > 0$  and  $\gamma > r - 1$  (Jorba and Simó 1992). Then, it is not difficult to check that the change of variables  $x = (I + \varepsilon P_1(\omega t))y$  transforms (4) into

$$\begin{aligned} \dot{y} &= \left( A_0 + \varepsilon^2 (I + \varepsilon P_1(\theta))^{-1} \tilde{Q}(\theta) P_1(\theta) \right) y, \\ \dot{\theta} &= \omega. \end{aligned} \quad (4)$$

Since this equation is similar to (1) but with  $\varepsilon^2$  instead of  $\varepsilon$ , the inductive scheme seems clear: average the quasi-periodic part of (4) and restart this process to obtain a sequence of matrices  $\{P_j\}_j$  and  $\{A_j\}_j$  such that the products of the matrices  $(I + \varepsilon^{2^{j-1}} P_j)$  converge to some (smooth and regular) matrix  $P$  and the matrices  $A_j$  converge to some matrix  $B$ . This convergence is studied in Bogoljubov et al. (1976) and Jorba and Simó (1992). Here it is important to note that this process can be carried out by a computer program to obtain the matrices  $P_1, P_2, \dots$ , and  $A_0, A_1, \dots$ , so that it is possible to use, for a value of  $j$  sufficiently large,  $P_j$  and  $A_j$  as good approximations to  $P$  and  $B$ , respectively. A discussion of the effectivity of this procedure is contained in Jorba et al. (1997) where, as an example, the following equation is considered:

$$\ddot{x} + (1 + \varepsilon q(t))x = 0,$$

where  $q(t) = \cos(\omega_1 t) + \cos(\omega_2 t)$ , with  $(\omega_1, \omega_2) = (\sqrt{2}, \sqrt{3})$ . The previous algorithm has been implemented as a C program for a given (and fixed) value of  $\varepsilon$ . The program computes and performs a finite number of changes of variables. To simplify and make the program more efficient, the functions involved are expanded as (truncated) Fourier series and their coefficients have been stored as double-precision variables. During the operations, coefficients less than  $10^{-20}$  have been dropped in order to control the size of the Fourier series that appears during the process. Of course, this introduces some (small)

numerical error in the results. After four changes of variables, the original equation becomes

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \left[ \begin{pmatrix} 0.0 & b_{12} \\ b_{21} & 0.0 \end{pmatrix} + R(t) \right] \begin{pmatrix} x \\ y \end{pmatrix},$$

where  $b_{12} = 1.000000366251255$  and  $b_{21} = -0.992421151834871$ . The remainder  $R$  is very small: the largest coefficient it contains is below  $10^{-16}$ . Note that the accuracy (relative error) of this remainder  $R$  is very poor due to the use of double precision arithmetic (15–16 digits) for the coefficients. See Jorba et al. (1997) for more details.

## The Solar System

Many physical systems are Hamiltonian, and they can be written in a perturbative way. That is, the Hamiltonian is written in the form  $H = H_0 + H_1$ , where  $H_0$  is integrable (i.e., solvable by closed formulas) and  $H_1$  is small. A paradigmatic example is the Solar system, where  $H_0$  describes the motion of  $N$  particles (the planets), each one being attracted by a fixed central mass (the Sun). Then, the dynamics of the planets is integrable and follows the Kepler laws. In this case,  $H_1$  contains the interactions between planets and that the Sun is not fixed at the center. Perturbation theory was initially developed to deal with this situation (see the textbooks by Moulton (1914), Brouwer and Clemence (1961), and Roy (2004)) and it has been extended to other fields.

The arrival of the electronic computer opened a huge range of new methods to study physical models, and some of them are based on taking advantage of ideas from the classical perturbation theory. For instance, there are several numerical studies of the long-term dynamics of the Solar system that exploit this perturbative structure in order to be more efficient both in computer speed and accuracy. There is a limitation of the total integration time that comes from the chaoticity of the system (Laskar 1989) but the results have a high interest for paleoclimatology: the variations of the Earth orbital elements induce some changes in the Earth climate that have an impact in the sedimentary records. Therefore, it is possible to

use the knowledge of the Solar system dynamics for the calibration of geological time scales by correlating the variation of orbital and rotational elements of the Earth with the existing geological records (Westerhold et al. 2008; Hoang et al. 2021).

An option to simplify (and to speed up) the integration of these equations is to use the ideas of perturbation theory to prepare the equations before the numerical integration. One of the main ideas is that the planetary motion combines quantities (angles) that move fast with quantities with a slow dynamics. A first tool to deal with fast angles is the averaging method. The idea is that these rapidly oscillating terms tend to average out without contributing to the long-term evolution, while more slowly varying resonant or secular terms accumulate to give significant contributions. In this direction, a set of very relevant works are (Laskar 1985, 1986, 1988) (see also references therein) where a sequence of symbolic manipulations on the equations of motion (that have been expanded in a Taylor-Fourier series in suitable variables) has been done to carry out a normalization procedure (with truncation up to suitable order) to produce a much better form for the differential equations which allows for a fast and efficient numerical integration, see also Laskar (1990, 1994). These formal computations are done by a specific computer algebra package, see Gastineau and Laskar (2011) for its last version.

A different way of exploiting the perturbative structure is by means of direct numerical integration methods that take advantage of it. A very simple (but very effective) idea is to use the classical representation of the Hamiltonian seen before,  $H = H_0 + H_1$ , where  $H_0$  contains the uncoupled two body interactions Sun-planet (so it is an integrable Hamiltonian) and  $H_1$  contains the interactions between planets (so it is small). This representation is exploited by Wisdom and Holman (1991) to develop a specific numerical integrator for the planetary problem that preserves the Hamiltonian structure. This idea of writing the Hamiltonian of the Solar system as the sum of these two parts has been used to construct other families of symplectic numerical integrators (McLachlan

1995; Chambers and Murison 2000; Laskar and Robutel 2001; Blanes et al. 2013).

## Dynamics Near an Equilibrium Point of a Hamiltonian System

The study of the motion near an equilibrium point of a Hamiltonian system is a very relevant topic of classical mechanics. Without loss of generality it is assumed that the equilibrium point is at the origin, and that the Hamiltonian system  $H$  is analytic, with  $\ell$  degrees of freedom, that can be expanded at the origin as  $H = H_2 + H_{>2}$  where  $H_2$  is a homogeneous polynomial of degree 2 and  $H_{>2}$  contains the higher order terms (the terms of degree 1 vanish since the origin is an equilibrium point). This is a perturbative setting since, near the origin,  $H_{>2}$  is much smaller than  $H_2$ . It is now very convenient to take advantage of this situation to obtain a quantitative description of the dynamics near the origin. Let us distinguish several cases.

The first case is when the equilibrium point is of elliptic type. This means that, in suitable variables,  $H_2$  can be written as

$$H_2(q, p) = \sum_{j=1}^{\ell} \frac{\omega_j}{2} (q_j^2 + p_j^2),$$

where  $\omega = (\omega_1, \dots, \omega_{\ell})$  are the frequencies of the linearized equations at the origin. Let us now assume that this is the non-resonant case, that is,  $\langle k, \omega \rangle \neq 0$  for all  $k \in \mathbb{Z}^{\ell} \setminus \{0\}$ , and that the Hamiltonian is expanded as  $H = H_2(q, p) + H_3(q, p) + \dots + H_N(q, p) + R_{N+1}(q, p)$ , where  $H_j$  is a homogeneous polynomial of degree  $j$  and  $R_{N+1}$  is the remainder (of order  $N+1$ ). An option in this situation is to start a normalizing procedure to simplify, order by order, the power expansion of the Hamiltonian up to degree  $N$ , to produce the so-called Birkhoff normal form (Meyer and Offin 2017). This normal form allows to easily introduce action-angle variables  $(I, \varphi)$  ( $I \in \mathbb{R}^{\ell}$ ,  $\varphi \in \mathbb{T}^{\ell}$ ) such that the Hamiltonian takes the form  $H = H_0(I) +$

$H_1(I, \varphi)$ , where  $H_0$  is a polynomial of degree  $N/2$  and  $H_1$  contains the non-normalized part (and it starts at order  $N/2 + 1$ ). Generically speaking, these normalizing transformations are not convergent and, when  $N$  goes to infinity, the normal form is seen as the power series of a  $C^\infty$  function. If the normalizing process is stop at a given  $N$ , then, as the remainder  $H_1$  is of order  $N/2 + 1$ , it turns out that in a small enough neighborhood of the origin this remainder is sufficiently small so that  $H_0$  gives an accurate description of the dynamics. The change of variables that relate the original coordinates with the final action-angle variables can be obtained in a similar way, so orbits in the normal form can be sent to original coordinates if required. Note that the dynamics given by  $H_0$  is quasi-periodic and this implies that, if the small term  $H_1$  is discarded, the origin is stable. For long-term dynamics, this remainder can be relevant and there are procedures to bound its size and, hence, to bound the speed of the possible instability (usually called Arnol'd diffusion (Arnol'd 1964)) that may take place around the point (Giorgilli et al. 1989; Simó 1989; Giorgilli and Skokos 1997; Benettin et al. 1998; Páez and Efthymiopoulos 2018; Caracciolo and Locatelli 2020). A similar machinery can be used for conservative maps (Simó and Vieira 2009).

An alternative to the computation of the Birkhoff normal form to bound Arnol'd diffusion is to compute formal first integrals (Marchal 1980), again as a perturbative series that starts with the second degree terms of the Hamiltonian ( $H_2$  above). Bounding the drift of a suitable truncation of these first integrals provides a bound on the diffusion speed Giorgilli (1979, 1988); Celletti and Giorgilli (1991).

It is important to note that, as this procedure is based on the manipulation of homogeneous polynomials, it can be carried out by a computer program. In fact, there is a long tradition in Celestial Mechanics of building algebraic manipulators to deal with these expansions (Broucke and Garthwaite 1969; Ricklefs et al. 1983; Broucke 1988; Meyer and Schmidt 1986; Brumberg et al. 1989; Jorba 1999). In many cases, floating point

coefficients are used for the coefficients of these expansions, but it is possible to use interval arithmetic for them (Jorba 1999) in order to control the rounding error of the operations, which can allow to produce a computer-assisted proof either to check the non-degeneracy condition of the KAM theorem (Meyer and Offin 2017) and to prove the existence of invariant tori, or to bound the diffusion speed.

Another interesting situation is when the equilibrium point has some unstable directions. This is what happens near the collinear points of the restricted three-body problem (RTBP for short, see Meyer and Offin (2017)) or in some problems of reaction dynamics in chemical physics (Uzer et al. 2002). The neighborhoods of the collinear points of the RTBP are interesting for space missions, since they are a suitable place for some space missions. In these cases, it is necessary to find trajectories that stay near these points and this can be done by the Lindstedt-Poincaré method which is a recursive procedure to find perturbative expansions for periodic and quasi-periodic orbits near a known solution, which can be implemented in a computer program. Using floating point numbers as coefficients of the expansion simplifies the coding and also produces a very efficient algorithm (Richardson 1980; Gómez et al. 1985, 1997, 1998; Jorba and Masdemont 1999). Another option to study the dynamics is the so-called reduction to the center manifold (Gómez et al. 1991), which is a partial normal form process that uncouples the hyperbolic directions from the elliptic ones, and that it allows to restrict the Hamiltonian to these elliptic directions to visualize the dynamics (Jorba and Masdemont 1999). In some situations, a complete normal form is preferred, for instance to analyze passages near the hyperbolic point (Duarte and Jorba 2021). When studying the reaction rate of chemical reactions, similar calculations appear (Jaffé et al. 2005; Haller et al. 2011). As it has been mentioned before, these numerical computations can be carried out using interval arithmetic so that the result can be used to produce a computer-assisted proof (Capiński and Roldán 2012).

A difficulty of the previous approaches is when the interesting region is beyond the radius of convergence of the previous expressions. An option is to use different sets of coordinates (Skokos and Dokoumetzidis 2000; Páez and Locatelli 2015).

## Time-Dependent Perturbations

There are several physical models that are written as an autonomous (usually dominant) part plus a time-dependent (usually small) part. The goal is to study the effect of this time-dependent part on the dynamics. A typical question is how this perturbation affects the stability. Well-known problems of this kind are the effective stability of the triangular points in the Elliptic RTBP, which can be seen as an autonomous part plus a time-dependent part coming from the eccentricity of the primaries. A standard tool is the computation of suitable normal forms, which is quite similar to the autonomous case (Jorba and Simó 1994; Lhotka et al. 2008). Another important modification of the RTBP is the Bicircular Problem (BCP; Huang (1960); Cronin et al. (1964)) and the Quasi-Bicircular Problem (QBCP; Andreu (2002)). These are periodically time-dependent perturbations of the Earth-Moon Circular RTBP that take into account the direct effect of the Sun on the infinitesimal particle. In these models, the equilibrium points are replaced by periodic orbits, and the dynamics around these periodic orbits can also be studied by means of suitable normalization procedures around periodic orbits (Simó et al. 1995; Jorba et al. 2020). As it happens in the autonomous case, the stability can also be studied by computing approximate first integrals, which now depend on time in a periodic way, see Gabern and Jorba (2001) for an example in the Sun-Jupiter-Saturn case. It is interesting to note that, under generic conditions of non-degeneracy and non-resonance Jorba and Villanueva (1997), a periodic orbit of the autonomous system becomes quasi-periodic when perturbed periodically, with two frequencies: the one of the periodic orbit plus the one of the perturbation. These quasi-periodic orbits can also be computed taking advantage of the perturbative situation (Castellà and Jorba 2000).

There are some physical models that are written as time-dependent quasi-periodic perturbations of autonomous models (Gómez et al. 2002). In these models, the simplest solution is a quasiperiodic orbit, that can be computed by means of suitable numerical methods (Castellà and Jorba 2000; Jorba and Olmedo 2009). Its linear stability can be analyzed by the method discussed in the Introduction, or by means of direct numerical methods (Jorba 2001). Then, it is possible to start a normalizing procedure to have a nonlinear description of the surrounding trajectories (Gabern and Jorba 2005), or to compute approximate first integrals. There are some models for chemical reactions that also include time-dependent perturbations (Zhang and de la Llave 2018).

The dynamics of these quasi-periodic models is more complex due to the high number of resonances. Therefore, they are very difficult to understand: as the model usually includes different effects, it can be difficult to find the relevance of each of them on the final properties of the system. The perturbative approach is to use a sequence of increasingly accurate (and increasingly complicated) models. First, the simplest model is studied in detail, and then this information is used to study the next model. This involves continuation methods, bifurcation analysis, normal forms, etc. Again, the result of this study is taken as starting point for the next model. At each stage new phenomena appear. At the end, therefore, it is possible to find (and understand) some phenomena present in the most realistic model but not present (even in a qualitative form) in the most simplified one. Examples of works using this approach are Gómez et al. (1985, 1987, 1991, 1993) and Jorba (2000).

Finally, there are also some academic examples of quasi-periodic time-dependent perturbations of autonomous systems. For a very accurate numerical perturbative treatment of a quasi-periodically forced pendulum, see Simó (1994).

## Quasi-Periodic Motions and KAM Theory

In many situations, KAM methods are constructive and, in some situations, they can be used as a



computational tool (Guzzo et al. 2020). For instance, it is possible to use the original Kolmogorov scheme (Kolmogorov 1954) to construct invariant tori near some Trojan asteroid in the Circular RTBP and to use this construction to show the effective stability of the asteroid (Gabern et al. 2005). The explicit construction of a normal form around a periodic orbit of an autonomous system is described in Jorba and Villanueva (1998).

On the other hand, rigorous numerical methods can be used to validate the hypotheses required for a suitable version of a theorem and then to show the existence of plenty of quasi-periodic motions near a known periodic orbit (Kapela and Simó 2017). These ideas are quite general and are being applied to many situations. For instance, to study the existence of quasi-periodic motions in realistic celestial mechanics problems (Robutel 1995) like the Sun-Jupiter-Saturn system (Locatelli and Giorgilli 2000, 2005, 2007) or the Sun-Jupiter-Saturn-Uranus system (Sansottera et al. 2011). These normalizing techniques can also be used to prove the stability for very long times (Giorgilli and Locatelli 2009; Martínez and Simó 2013; Giorgilli et al. 2017).

A problem of great practical interest is the motion of particles around the Earth. This, of course, includes artificial satellites but also particles of space debris whose dynamics is becoming very important due to the danger they pose for operating satellites and spacecrafts. Their motion is close to Keplerian, and the main perturbations come from the non-symmetric mass distribution of the Earth, the gravitational effects of Sun and Moon and the Solar radiation pressure. These orbits can be studied numerically but, since they are close to Keplerian motion, they can also be studied using a combination of numerical and perturbation methods (Lara et al. 2012, 2014, 2018, 2020; Celletti and Galeš 2014, 2018; Celletti et al. 2020; Gachet et al. 2017).

More recently, a new point of view has appeared. Assume that, by means of numerical methods, we have been able to obtain an accurate approximation to a periodic or quasi-periodic orbit of a Hamiltonian system  $H$ , and the goal is to show that there exists a true periodic or quasi-

periodic orbit nearby. The idea is to think that the result of the computation is a true orbit of a nearby Hamiltonian system,  $\hat{H}$ , and then write  $H$  as  $\hat{H} + (H - \hat{H})$ , where  $H - \hat{H}$  should be small. This is again a perturbative situation, of a very special kind. Under generic assumptions it is possible to prove that, if the computation of the initial periodic or quasi-periodic orbit is sufficiently accurate (this is like saying that  $H - \hat{H}$  is small enough), then there is a true periodic or quasi-periodic orbit nearby (de la Llave et al. 2005). An advantage of this approach is that the hypotheses can be checked on the initial Hamiltonian (i.e., no need to use action-angle variables). This kind of results are key ingredients to produce computer-assisted proofs and to validate numerical computations (that is, to have a true bound on the error of the numerical computation). This allows to be confident of the results in some difficult situations when it is nearly impossible to have a reliable control of the errors by other means. For examples, see Figueras and Haro (2012) and Figueras et al. (2017).

## The Parametrization Method

This method was used in the 1980s to compute invariant manifolds by C. Simó (see also Franceschini and Russo (1981)), but it is also a very good tool to prove their existence as shown by Cabré et al. (2003a, b, 2005). A very good exposition of the method can be found in the book by Haro et al. (2016).

The idea of the method is to write the manifold of an equilibrium point (or fixed point for a discrete system) in parametric form, to expand it in power series, and to find recursively the coefficients of this expansion by imposing that the manifold is invariant under the dynamics (Simó 1990). Under suitable hypotheses, this power series is convergent in a neighborhood of the point giving, at the same time, a computational algorithm and the scheme of a proof. These ideas can be extended for manifolds around periodic orbits or invariant tori (Haro and de la Llave 2006a, b; Castelli et al. 2015).



A technique that can be combined with the parametrization method is the so-called jet transport. Jet transport is based on using automatic differentiation (Griewank 2000) on the code of a numerical method for initial value problems of ODEs. The main idea is to replace the basic arithmetic by an arithmetic of (truncated) formal power series in several variables. Note that a formal power series contains the value of a function in a point (the constant term) and their derivatives at this point (the coefficients of each monomial), and the propagation of these power series through the operations of the numerical integration gives the derivative of the final point with respect to the initial data and/or parameters (Berz and Makino 1998; Alessi et al. 2008). This implies that it is possible to compute derivatives of any order of the flow and, therefore, derivatives of the flow. In particular, this allows to use the parametrization method on a Poincaré section (Jorba and Nicolás 2021).

## Other Situations

This section is devoted to survey some results for general dynamical systems. A first important context is given by delay differential equations. Since the phase space is of infinite dimension, it is very helpful to take advantage of perturbative schemes to avoid brute-force computations. Examples of finding perturbative expansions in some concrete situations are Groothedde and Mireles-James (2017), Yang et al. (2021), and Gimeno et al. (2021). Moreover, some of these techniques can also be used for some partial differential equations (Zgliczynski and Mischaikow 2001; Arioli and Koch 2010; Breden et al. 2013; Reinhardt and Mireles-James 2019).

Perturbative techniques can also be adapted to do computations on dissipative systems (Calleja and Celletti 2010). Among them, there are conformally symplectic systems which are similar to conservative systems, in the sense that the symplectic structure is not preserved but multiplied by a constant. This makes that they can be studied by similar methods (Calleja et al. 2013, 2021; Bustamante and Calleja 2019, 2021).

There are works using perturbative techniques on more general dynamical systems, like computing expansions of invariant manifolds or studying chaos and diffusion (Lessard et al. 2014; Castelli et al. 2015; van den Berg et al. 2016; Gelfreich and Vieiro 2018; Meiss et al. 2018).

## Future Directions

Computational methods in perturbation theory are experiencing a big development in the recent years, combined with new theoretical results. In particular, very relevant advances are taking place in some computational methods like the parametrization method and the jet transport technique. They are allowing to develop high-order perturbative expansions in new situations. These new techniques are expected to have an impact in the design of space missions during the next years (Chen et al. 2020; Kumar et al. 2021). Another problem of high interest in astrodynamics is the dynamics of space debris (Celletti et al. 2017), which has become very important for space navigation. At present, there is a lot of research going on in this direction so it is expected to have important advances in the forthcoming years.

A field in which there is a lot of activity is the production of computer-assisted proofs (CAP). They allow to produce formal mathematical proofs in situations that are far beyond the reach of the classical mathematical machinery (van den Berg and Jaquette 2018; González et al. 2021; Valvo and Locatelli 2021). It is expected that this research line will continue growing. In this direction, the parametrization method is being extended to more situations, giving rise to new methods and applications (Haro and Mondelo 2021).

**Acknowledgments** This work has been supported by the Spanish grant PGC2018-100699-B-I00 (MCIU/AEI/FEDER, UE) and the Catalan grant 2017 SGR 1374. The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 734557.

## Bibliography

- Alessi E, Farrés A, Jorba A, Simó C, Vieira A (2008) Efficient usage of self validated integrators for space applications. Ariadna final report, contract no. 20783/07/nl/cb, ESTEC (European Space Agency)
- Andreu M (2002) Dynamics in the center manifold around  $L_2$  in the quasi-bicircular problem. *Celest Mech Dyn Astron* 84(2):105–133
- Arioli G, Koch H (2010) Computer-assisted methods for the study of stationary solutions in dissipative systems, applied to the Kuramoto-Sivashinski equation. *Arch Ration Mech Anal* 197:1033–1051. <https://doi.org/10.1007/s00205-010-0309-7>
- Arnol'd V (1964) Instability of dynamical systems with several degrees of freedom. *Soviet Math Dokl* 5: 581–585
- Benettin G, Fassò F, Guzzo M (1998) Nekhoroshev-stability of  $L_4$  and  $L_5$  in the spatial restricted three-body problem. *Regul Chaotic Dyn* 3(3):56–72
- Berz M, Makino K (1998) Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliab Comput* 4:361–369. <https://doi.org/10.1023/A:1024467732637>
- Blanes S, Casas F, Farrés A, Laskar J, Makazaga J, Murua A (2013) New families of symplectic splitting methods for numerical integration in dynamical astronomy. *Appl Numer Math* 68:58–72. <https://doi.org/10.1016/j.apnum.2013.01.003>
- Bogoljubov NN, Mitropoliskii JA, A. M. Samoilenko. (1976) *Methods of accelerated convergence in nonlinear mechanics*. Springer
- Breden M, Lessard J, Vanicat M (2013) Global bifurcation diagrams of steady states of systems of PDEs via rigorous numerics: a 3-component reaction-diffusion system. *Acta Appl Math* 128:113–152. <https://doi.org/10.1007/s10440-013-9823-6>
- Broucke R (1988) A Fortran-based poisson series processor and its applications in celestial mechanics. *Celestial Mech* 45:255–265. <https://doi.org/10.1007/BF01229008>
- Broucke R, Garthwaite K (1969) A programming system for analytical series expansions on a computer. *Celestial Mech* 1:271–284
- Brouwer D, Clemence G (1961) *Methods of celestial mechanics*. Academic, New York/London
- Brumberg V, Tarasevich S, Vasiliev N (1989) Specialized celestial mechanics systems for symbolic manipulation. *Celestial Mech* 45:149–162
- Bustamante A, Calleja R (2019) Computation of domains of analyticity for the dissipative standard map in the limit of small dissipation. *Phys D* 395:15–23. <https://doi.org/10.1016/j.physd.2019.02.006>
- Bustamante A, Calleja R (2021) Corrigendum and addendum to “computation of domains of analyticity for the dissipative standard map in the limit of small dissipation”. [*Phys D* 395:15–23 (2019)]. *Phys D* 417:132837. <https://doi.org/10.1016/j.physd.2020.132837>
- Cabrè X, Fontich E, de la Llave R (2003a) The parameterization method for invariant manifolds. I. Manifolds associated to non-resonant subspaces. *Indiana Univ Math J* 52(2):283–328. <https://doi.org/10.1512/iumj.2003.52.2245>. ISSN 0022-2518
- Cabrè X, Fontich E, de la Llave R (2003b) The parameterization method for invariant manifolds. II. Regularity with respect to parameters. *Indiana Univ Math J* 52(2): 329–360. <https://doi.org/10.1512/iumj.2003.52.2407>. ISSN 0022-2518
- Cabrè X, Fontich E, de la Llave R (2005) The parameterization method for invariant manifolds. III. Overview and applications. *J Differential Equations* 218(2): 444–515. <https://doi.org/10.1016/j.jde.2004.12.003>. ISSN 0022-0396
- Calleja R, Celletti A (2010) Breakdown of invariant attractors for the dissipative standard map. *Chaos* 20(013121). <https://doi.org/10.1063/1.3335408>
- Calleja R, Celletti A, de la Llave R (2013) A KAM theory for conformally symplectic systems: efficient algorithms and their validation. *J Differential Equations* 255(5): 978–1049. <https://doi.org/10.1016/j.jde.2013.05.001>
- Calleja R, Canadell M, Haro A (2021) Non-twist invariant circles in conformally symplectic systems. *Commun Nonlinear Sci Numer Simul* 96:105695. <https://doi.org/10.1016/j.cnsns.2021.105695>
- Capiński M, Roldán P (2012) Existence of a center manifold in a practical domain around  $L_1$  in the restricted three-body problem. *SIAM J Appl Dyn Syst* 11(1): 285–318
- Caracciolo C, Locatelli U (2020) Computer-assisted estimates for Birkhoff normal forms. *J Comput Dyn* 7(2): 425–460
- Castellà E, Jorba A (2000) On the vertical families of two-dimensional tori near the triangular points of the Bicircular problem. *Celest Mech Dyn Astron* 76(1): 35–54
- Castelli R, Lessard J-P, Mireles James JD (2015) Parameterization of invariant manifolds for periodic orbits I: efficient numerics via the Floquet normal form. *SIAM J Appl Dyn Syst* 14(1):132–167. <https://doi.org/10.1137/140960207>. ISSN 1536-0040
- Celletti A (2010) *Stability and chaos in celestial mechanics*. Springer praxis books. Springer, Berlin/Heidelberg. <https://doi.org/10.1007/978-3-540-85146-2>
- Celletti A, Chierchia L (1998) KAM stability estimates in celestial mechanics. *Planet Space Sci* 46(11–12): 1433–1440. [https://doi.org/10.1016/S0032-0633\(97\)00219-5](https://doi.org/10.1016/S0032-0633(97)00219-5)
- Celletti A, Galeš C (2014) On the dynamics of space debris: 1:1 and 2:1 resonances. *J Nonlinear Sci* 24:1231–1262. <https://doi.org/10.1007/s00332-014-9217-6>
- Celletti A, Galeš C (2018) Dynamics of resonances and equilibria of low earth objects. *SIAM J Appl Dyn Syst* 17(1):203–235. <https://doi.org/10.1137/17M118671>
- Celletti A, Giorgilli A (1991) On the stability of the Lagrangian points in the spatial restricted three body problem. *Celest Mech Dyn Astron* 50(1):31–58

- Celletti A, Efthymiopoulos C, Gachet F, Galeš C, Pucacco G (2017) Dynamical models and the onset of chaos in space debris. *Int J Non Linear Mech* 90:147–163. <https://doi.org/10.1016/j.ijnonlinmec.2016.12.015>. ISSN 0020-7462
- Celletti A, Galeš C, Lhotka C (2020) Resonances in the Earth's space environment. *Commun Nonlinear Sci Numer Simul* 84:105185. <https://doi.org/10.1016/j.cnsns.2020.105185>
- Chambers J, Murison M (2000) Pseudo-high-order symplectic integrators. *Astron J* 119(1):425–433. <https://doi.org/10.1086/301161>
- Chen J, Masdemont J, Gómez G, Yuan J (2020) Analysis of jet transport-based geostationary trajectory uncertainty propagation. *J Guid Control Dyn* 43(6):1210–1219. <https://doi.org/10.2514/1.G004110>
- Cronin J, Richards P, Russell L (1964) Some periodic solutions of a four-body problem. *Icarus* 3:423–428
- de la Llave R, González A, Jorba À, Villanueva J (2005) KAM theory without action-angle variables. *Nonlinearity* 18(2):855–895
- Duarte G, Jorba À (2021) Using normal forms to study Oterma's transition in the planar RTBP. Preprint
- Farrés A, Laskar J, Blanes S, Casas F, Makazaga J, Murua A (2013) High precision symplectic integrators for the solar system. *Celest Mech Dyn Astron* 116:141–174. <https://doi.org/10.1007/s10569-013-9479-6>
- Féjoz J, Guardia M, Kaloshin V, Roldán P (2016) Kirkwood gaps and diffusion along mean motion resonances in the restricted planar three-body problem. *J Eur Math Soc* 18(10):2315–2403. <https://doi.org/10.4171/JEMS/642>
- Figueras J-L, Haro A (2012) Reliable computation of robust response tori on the verge of break-down. *SIAM J Appl Dyn Syst* 11(2):597–628
- Figueras J-L, Haro A, Luque A (2017) A rigorous computer-assisted application of KAM theory: a modern approach. *Found Comput Math* 17:1123–1193. <https://doi.org/10.1007/s10208-016-9339-3>
- Franceschini V, Russo L (1981) Stable and unstable manifolds of the Hénon mapping. *J Stat Phys* 25(4):757–769
- Gabern F, Jorba À (2001) A restricted four-body model for the dynamics near the Lagrangian points of the Sun-Jupiter system. *Discrete Contin Dyn Syst Ser B* 1(2):143–182
- Gabern F, Jorba À (2005) Effective computation of the dynamics around a two-dimensional torus of a Hamiltonian system. *J Nonlinear Sci* 15(3):159–182
- Gabern F, Jorba À, Locatelli U (2005) On the construction of the Kolmogorov normal form for the Trojan asteroids. *Nonlinearity* 18(4):1705–1734
- Gachet F, Celletti A, Pucacco G, Efthymiopoulos C (2017) Geostationary secular dynamics revisited: application to high area-to-mass ratio objects. *Celest Mech Dyn Astron* 128:149–181. <https://doi.org/10.1007/s10569-016-9746-4>
- Gastineau M, Laskar J (2011) TRIP: A computer algebra system dedicated to celestial mechanics and perturbation series. *ACM Commun Comput Algebra* 44(3/4):194–197. <https://doi.org/10.1145/1940475.1940518>. ISSN 1932-2240
- Gelfreich V, Vieira A (2018) Interpolating vector fields for near identity maps and averaging. *Nonlinearity* 31(9):4263–4289. <https://doi.org/10.1088/1361-6544/aacb8e>. ISSN 0951-7715
- Gimeno J, Yang J, de la Llave R (2021) Numerical computation of periodic orbits and isochrones for state-dependent delay perturbation of an ODE in the plane. *SIAM J Appl Dyn Syst* 20(3):1514–1543. <https://doi.org/10.1137/20M1336965>
- Giorgilli A (1979) A computer program for integrals of motion. *Comp Phys Comm* 16:331–343
- Giorgilli A (1988) Rigorous results on the power expansions for the integrals of a Hamiltonian system near an elliptic equilibrium point. *Ann Inst H Poincaré A* 48(4):423–439
- Giorgilli A (1995) Quantitative methods in classical perturbation theory. In: Roy A, Steves B (eds) From Newton to chaos, volume 336 of NATO ASI series (series B: physics). Springer, Boston. [https://doi.org/10.1007/978-1-4899-1085-1\\_3](https://doi.org/10.1007/978-1-4899-1085-1_3)
- Giorgilli A, Locatelli U (2006) Canonical perturbation theory for nearly integrable systems. In: Steves B, Maciejewski A, Hendry M (eds) Chaotic worlds: from order to disorder in gravitational N-body dynamical systems. Springer. [https://doi.org/10.1007/978-1-4020-4706-0\\_1](https://doi.org/10.1007/978-1-4020-4706-0_1)
- Giorgilli A, Locatelli U (2009) Kolmogorov and Nekhoroshev theory for the problem of three bodies. *Celest Mech Dyn Astron* 104:159–173. <https://doi.org/10.1007/s10569-009-9192-7>
- Giorgilli A, Sansottera M (2012) Methods of algebraic manipulation in perturbation theory. In: Cincotta P, Giordano C, Efthymiopoulos C (eds) Proceedings of the third la plata international school on astronomy and geophysics. Universidad Nacional de La Plata and Asociación Argentina de Astronomía
- Giorgilli A, Skokos C (1997) On the stability of the Trojan asteroids. *Astron Astrophys* 317:254–261
- Giorgilli A, Delshams A, Fontich E, Galgani L, Simó C (1989) Effective stability for a Hamiltonian system near an elliptic equilibrium point, with an application to the restricted three body problem. *J Differential Equations* 77:167–198
- Giorgilli A, Locatelli U, Sansottera M (2017) Secular dynamics of a planar model of the Sun-Jupiter-Saturn-Uranus system; effective stability in the light of Kolmogorov and Nekhoroshev theories. *Regul Chaotic Dyn* 22:54–77. <https://doi.org/10.1134/S156035471701004X>
- Gómez G, Llibre J, Martínez R, Simó C (1985) Station keeping of libration point orbits. ESOC contract 5648/83/D/JS(SC), final report, European Space Agency. Reprinted as Dynamics and mission design near libration points. Vol. I, Fundamentals: the case of collinear libration points, volume 2 of World scientific monograph series in mathematics, 2001

- Gómez G, Llibre J, Martínez R, Simó C (1987) Study on orbits near the triangular libration points in the perturbed Restricted Three-Body Problem. ESOC contract 6139/84/D/JS(SC), final report, European Space Agency. Reprinted as Dynamics and mission design near libration points. Vol. II, Fundamentals: the case of triangular libration points, volume 3 of World scientific monograph series in mathematics, 2001
- Gómez G, Jorba À, Masdemont J, Simó C (1991) Study refinement of semi-analytical Halo orbit theory. ESOC contract 8625/89/D/MD(SC), final report, European Space Agency. Reprinted as Dynamics and mission design near libration points. Vol. III, Advanced methods for collinear points, volume 4 of World scientific monograph series in mathematics, 2001
- Gómez G, Jorba À, Masdemont J, Simó C (1993) Study of Poincaré maps for orbits near Lagrangian points. ESOC contract 9711/91/D/IM(SC), final report, European Space Agency. Reprinted as Dynamics and mission design near libration points. Vol. IV, Advanced methods for triangular points, volume 5 of World scientific monograph series in mathematics, 2001
- Gómez G, Masdemont J, S. S. (1997) Lissajous orbits around halo orbits. *Adv Astronaut Sci* 95:117–134
- Gómez G, Masdemont J, Simó C (1998) Quasihalo orbits associated with libration points. *J Astronaut Sci* 46(2): 135–176
- Gómez G, Masdemont J, Mondelo J (2002) Solar system models with a selected set of frequencies. *Astron Astrophys* 390(2):733–749
- González A, Haro A, de la Llave R (2021) Efficient and reliable algorithms for the computation of non-twist invariant circles. *Found Comput Math*. <https://doi.org/10.1007/s10208-021-09517-9>
- Griewank A (2000) Evaluating derivatives. SIAM, Philadelphia. ISBN 0-89871-451-6
- Groothedde C, Mireles-James J (2017) Parameterization method for unstable manifolds of delay differential equations. *J Comput Dyn* 4(1&2):21–70. <https://doi.org/10.3934/jcd.2017002>
- Guzzo M, Efthymiopoulos C, Paez R (2020) Semi-analytic computations of the speed of Arnold diffusion along single resonances in a priori stable Hamiltonian systems. *J Nonlinear Sci* 30:851–901. <https://doi.org/10.1007/s00332-019-09594-9>
- Hairer E, Lubich C, Wanner G (2006) Geometric numerical integration, volume 31 of Springer series in computational mathematics, 2nd rev edn. Springer. <https://doi.org/10.1007/3-540-30666-8>
- Haller G, Uzer T, Palacián J, Yanguas P, Jaffé C (2011) Transition state geometry near higher-rank saddles in phase space. *Nonlinearity* 24(2):527–561. <https://doi.org/10.1088/0951-7715/24/2/007>
- Haro A, de la Llave R (2006a) A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: rigorous results. *J Differential Equations* 228(2):530–579. <https://doi.org/10.1016/j.jde.2005.10.005>. ISSN 0022-0396
- Haro À, de la Llave R (2006b) A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: numerical algorithms. *Discrete Contin Dyn Syst Ser B* 6(6):1261–1300. <https://doi.org/10.3934/dcdsb.2006.6.1261>. ISSN 1531-3492
- Haro A, Mondelo J (2021) Flow map parameterization methods for invariant tori in Hamiltonian systems. *Commun Nonlinear Sci Numer Simul*, 101:Paper No. 105859, 34. <https://doi.org/10.1016/j.cnsns.2021.105859>. ISSN 1007-5704
- Haro A, Canadell M, Luque A, Mondelo J-M, Figueras J-L (2016) The parameterization method for invariant manifolds. From rigorous results to effective computations, volume 195 of Applied mathematical sciences. Springer
- Henrard J (1989) A survey of Poisson series processors. *Celest Mech Dyn Astron* 45:245–253
- Hoang N, Mogavero F, Laskar J (2021) Chaotic diffusion of the fundamental frequencies in the solar system. *Astron Astrophys* 654:A156. <https://doi.org/10.1051/0004-6361/202140989>
- Huang S (1960) Very restricted four-body problem. Technical note TN D-501, Goddard Space Flight Center, NASA. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19890068606.pdf>
- Jaffé C, Kawai S, Palacián J, Yanguas P, Uzer T (2005) A new look at the transition state: Wigner's dynamical perspective revisited. In: Geometric structures of phase space in multidimensional chaos. Wiley, pp 171–216. <https://doi.org/10.1002/0471712531.ch3>. Chapter 3. ISBN 9780471712534
- Jorba À (1999) A methodology for the numerical computation of normal forms, centre manifolds and first integrals of Hamiltonian systems. *Exp Math* 8(2):155–195
- Jorba À (2000) A numerical study on the existence of stable motions near the triangular points of the real Earth-Moon system. *Astron Astrophys* 364(1): 327–338
- Jorba À (2001) Numerical computation of the normal behaviour of invariant curves of  $n$ -dimensional maps. *Nonlinearity* 14(5):943–976
- Jorba À, Masdemont J (1999) Dynamics in the centre manifold of the collinear points of the restricted three body problem. *Phys D* 132:189–213
- Jorba À, Nicolás B (2021) Using invariant manifolds to capture an asteroid near the  $L_3$  point of the Earth-Moon bicircular model. *Commun Nonlinear Sci Numer Simul* 102:105948. <https://doi.org/10.1016/j.cnsns.2021.105948>
- Jorba À, Olmedo E (2009) On the computation of reducible invariant tori on a parallel computer. *SIAM J Appl Dyn Syst* 8(4):1382–1404
- Jorba À, Simó C (1992) On the reducibility of linear differential equations with quasiperiodic coefficients. *J Differential Equations* 98:111–124
- Jorba À, Simó C (1994) Effective stability for periodically perturbed Hamiltonian systems. In: Seimenis, pp 245–252. ISBN 0–306–44808–4



- Jorba À, Villanueva J (1997) On the persistence of lower dimensional invariant tori under quasi-periodic perturbations. *J Nonlinear Sci* 7:427–473
- Jorba À, Villanueva J (1998) Numerical computation of normal forms around some periodic orbits of the restricted three body problem. *Phys D* 114(3–4):197–229
- Jorba À, Ramírez-Ros R, Villanueva J (1997) Effective reducibility of quasiperiodic linear equations close to constant coefficients. *SIAM J Math Anal* 28(1): 178–188
- Jorba À, Jorba-Cuscó M, Rosales J (2020) The vicinity of the Earth-Moon  $L_1$  point in the Bicircular problem. *Celest Mech Dyn Astron* 132(11). <https://doi.org/10.1007/s10569-019-9940-2>
- Kapela T, Simó C (2017) Rigorous KAM results around arbitrary periodic orbits for Hamiltonian systems. *Nonlinearity* 30(3):965–986. <https://doi.org/10.1088/1361-6544/aa4ff3>
- Kolmogorov A (1954) On the persistence of conditionally periodic motions under a small change of the Hamilton function. *Dokl Acad Nauk SSSR* 98(4):527–530
- Kumar B, Anderson R, de la Llave R (2021) High-order resonant orbit manifold expansions for mission design in the planar circular restricted 3-body problem. *Commun Nonlinear Sci Numer Simul* 97:105691. <https://doi.org/10.1016/j.cnsns.2021.105691>
- Lara M, San-Juan J, López L (2012) Semianalytic integration of high-altitude orbits under lunisolar effects. *Math Probl Eng* 659396:2012. <https://doi.org/10.1155/2012/659396>
- Lara M, San-Juan J, López-Ochoa L (2014) Efficient semi-analytic integration of GNSS orbits under tesseral effects. *Acta Astron* 102:355–366. <https://doi.org/10.1016/j.actaastro.2013.11.006>
- Lara M, San-Juan J, Hautesserres D (2018) HEOSAT: a mean elements orbit propagator program for highly elliptical orbits. *CEAS Space J* 10:3–23. <https://doi.org/10.1007/s12567-017-0152-x>
- Lara M, López R, Pérez I, San-Juan J (2020) Exploring the long-term dynamics of perturbed Keplerian motion in high degree potential fields. *Commun Nonlinear Sci Numer Simul* 82:105053. <https://doi.org/10.1016/j.cnsns.2019.105053>
- Laskar J (1985) Accurate methods in general planetary theory. *Astron Astrophys* 144(1):133–146
- Laskar J (1986) Secular terms of classical planetary theories using the results of general theory. *Astron Astrophys* 157(1):59–70
- Laskar J (1988) Secular evolution of the solar system over 10 million years. *Astron Astrophys* 198(1–2):341–362
- Laskar J (1989) A numerical experiment on the chaotic behaviour of the solar system. *Nature* 338:237–238. <https://doi.org/10.1038/338237a0>
- Laskar J (1990) The chaotic motion of the solar system: a numerical estimate of the size of the chaotic zones. *Icarus* 88:266–291
- Laskar J (1994) Large-scale chaos in the solar system. *Astron Astrophys* 287:L9–L12
- Laskar J, Robutel P (2001) High order symplectic integrators for perturbed Hamiltonian systems. *Celest Mech Dyn Astron* 80:39–62
- Lessard J, Mireles-James J, Reinhardt C (2014) Computer assisted proof of transverse saddle-to-saddle connecting orbits for first order vector fields. *J Dyn Diff Equat* 26:267–313. <https://doi.org/10.1007/s10884-014-9367-0>
- Lhotka C, Efthymiopoulos C, Dvorak R (2008) Nekhoroshev stability at  $L_4$  or  $L_5$  in the elliptic-restricted three-body problem – application to Trojan asteroids. *Mon Not R Astron Soc* 384(3):1165–1177
- Locatelli U, Giorgilli A (2000) Invariant tori in the secular motions of the three-body planetary systems. *Celest Mech Dyn Astron* 78:47–74. <https://doi.org/10.1023/A:1011139523256>
- Locatelli U, Giorgilli A (2005) Construction of Kolmogorov’s normal form for a planetary system. *Regul Chaotic Dyn* 10(2):153–171. <https://doi.org/10.1070/RD2005v010n02ABEH000309>
- Locatelli U, Giorgilli A (2007) Invariant tori in the Sun-Jupiter-Saturn system. *Discrete Contin Dyn Syst Ser B* 7(2):377–398
- Marchal C (1980) The quasi integrals. *Celestial Mech* 21: 183–191
- Martínez R, Simó C (2013) On the stability of the Lagrangian homographic solutions in a curved three-body problem on  $\mathbb{S}^2$ . *Discrete Contin Dyn Syst* 33(3):1157–1175
- McLachlan R (1995) Composition methods in the presence of small parameters. *BIT Numer Math* 35:258–268. <https://doi.org/10.1007/BF01737165>
- Meiss J, Miguel N, Simó C, Viegro A (2018) Accelerator modes and anomalous diffusion in 3D volume-preserving maps. *Nonlinearity* 31(12):5615–5642. <https://doi.org/10.1088/1361-6544/aae69f>. ISSN 0951-7715
- Meyer K, Offin D (2017). ISBN 978-3-319-53690-3) Introduction to Hamiltonian dynamical systems and the N-body problem, volume 90 of Applied mathematical sciences, 3rd edn. Springer, New York. <https://doi.org/10.1007/978-3-319-53691-0>
- Meyer KR, Schmidt DS (1986) The stability of the Lagrange triangular point and a theorem of Arnold. *J Differential Equations* 62(2):222–236. [https://doi.org/10.1016/0022-0396\(86\)90098-7](https://doi.org/10.1016/0022-0396(86)90098-7)
- Morbidelli A (2002) Modern integrations of solar system dynamics. *Annu Rev Earth Planet Sci* 30(1):89–112. <https://doi.org/10.1146/annurev.earth.30.091201.140243>
- Moulton F (1914) An introduction to celestial mechanics, 2nd rev edn. The Macmillan. Reprinted by Dover Publications, 1984
- Páez R, Efthymiopoulos C (2018) Secondary resonances and the boundary of effective stability of Trojan motions. *Celest Mech Dyn Astron* 130(20). <https://doi.org/10.1007/s10569-017-9814-4>
- Páez R, Locatelli U (2015) Trojan dynamics well approximated by a new Hamiltonian normal form. *Mon Not*

- R Astron Soc 453(2):2177–2188. <https://doi.org/10.1093/mnras/stv1792>
- Reinhardt C, Mireles-James J (2019) Fourier-Taylor parameterization of unstable manifolds for parabolic partial differential equations formalism, implementation and rigorous validation. *Indag Math* 30(1):39–80. <https://doi.org/10.1016/j.indag.2018.08.003>
- Richardson D (1980) Analytic construction of periodic orbits about the collinear points. *Celest Mech* 22(3): 241–253
- Ricklefs R, Jefferys W, Broucke R (1983) A general pre-compiler for algebraic manipulation. *Celest Mech* 30: 179–190. <https://doi.org/10.1007/BF01232805>
- Robutel P (1995) Stability of the planetary three-body problem. *Celest Mech Dyn Astron* 62:219–261. <https://doi.org/10.1007/BF00692089>
- Roy A (2004) *Orbital motion*, 4th edn. CRC Press
- San-Juan F, Abad A (2001) Algebraic and symbolic manipulation of Poisson series. *J Symb Comput* 32(5): 565–572. <https://doi.org/10.1006/jsc.2000.0396>
- Sansottera M, Locatelli U, Giorgilli A (2011) A semi-analytic algorithm for constructing lower dimensional elliptic tori in planetary systems. *Celest Mech Dyn Astron* 111:337–361. <https://doi.org/10.1007/s10569-011-9375-x>
- Seimenis J (ed) (1994) *Hamiltonian mechanics: integrability and chaotic behaviour*, volume 331 of NATO Adv. Sci. Inst. Ser. B Phys. Held in Toruń, Poland, 28 June–2 July 1993. Plenum, New York. ISBN 0–306–44808–4
- Simó C (1989) Estabilitat de sistemes Hamiltonians. *Mem Real Acad Cienc Artes Barcelona* 48(7):303–348
- Simó C (1990) On the analytical and numerical approximation of invariant manifolds. In: Benest D, Froeschlé C (eds) *Modern methods in celestial mechanics*. Ed. Frontières, pp 285–330. Reprinted at <http://www.maia.ub.es/dsg/2004/index.html>
- Simó C (1994) Averaging under fast quasiperiodic forcing. In: Seimenis, pp 13–34. ISBN 0–306–44808–4
- Simó C (1996) Effective computations in Hamiltonian dynamics. In: *Mécanique céleste*, volume 1996 of SMF Journ. Annu. Soc. Math. France, Paris, p 23
- Simó C (1998) Effective computations in celestial mechanics and astrodynamics. In: Rumyantsev V, Karapetyan A (eds) *Modern methods of analytical mechanics and their applications*, volume 387 of CISM courses and lectures. Springer
- Simó C, Vieiro A (2009) Resonant zones, inner and outer splittings in generic and low order resonances of area preserving maps. *Nonlinearity* 22(5):1191–1245. <https://doi.org/10.1088/0951-7715/22/5/012>
- Simó C, Gómez G, Jorba A, Masdemont J (1995) The Bicircular model near the triangular libration points of the RTBP. In: Roy A, Steves B (eds) *From Newton to chaos*. Plenum Press, New York, pp 343–370
- Skokos C, Dokoumetzidis A (2000) Effective stability of the Trojan asteroids. *Astron Astrophys* 367:729–736
- Uzer T, Jaffé C, Palacián J, Yanguas P, Wiggins S (2002) The geometry of reaction dynamics. *Nonlinearity* 15(4):957–992. <https://doi.org/10.1088/0951-7715/15/4/301>
- Valvo L, Locatelli U (2021) Hamiltonian control of magnetic field lines: computer assisted results proving the existence of KAM barriers. Preprint
- van den Berg J, Jaquette J (2018) A proof of Wright’s conjecture. *J Differential Equations* 264(12): 7412–7462. <https://doi.org/10.1016/j.jde.2018.02.018>. ISSN 0022-0396
- van den Berg J, Mireles-James J, Reinhardt C (2016) Computing (un)stable manifolds with validated error bounds: non-resonant and resonant spectra. *J Nonlinear Sci* 26:1055–1095. <https://doi.org/10.1007/s00332-016-9298-5>
- Westerhold T, Röhl U, Raffi I, Fornaciari E, Monechi S, Reale V, Bowles J, Evans H (2008) Astronomical calibration of the Paleocene time. *Palaeogeogr Palaeoclimatol Palaeoecol* 257(4):377–403. <https://doi.org/10.1016/j.palaeo.2007.09.016>
- Wisdom J, Holman M (1991) Symplectic maps for the  $N$ -body problem. *Astron J* 102(4):1528–1538. <https://doi.org/10.1086/115978>
- Yang J, Gimeno J, de la Llave R (2021) Parameterization method for state-dependent delay perturbation of an ordinary differential equation. *SIAM J Math Anal* 53(4):4031–4067. <https://doi.org/10.1137/20M1311430>
- Zgliczynski P, Mischakow K (2001) Rigorous numerics for partial differential equations: the Kuramoto-Sivashinsky equation. *Found Comput Math* 1: 255–288. <https://doi.org/10.1007/s002080010010>
- Zhang L, de la Llave R (2018) Transition state theory with quasi-periodic forcing. *Commun Nonlinear Sci Numer Simul* 62:229–243. <https://doi.org/10.1016/j.cnsns.2018.02.014>



# Perturbation Analysis of Parametric Resonance

Ferdinand Verhulst

Mathematisch Instituut, University of Utrecht,  
Utrecht, The Netherlands

## Article Outline

Glossary

Definition of the Subject

Introduction

Perturbation Techniques

Parametric Excitation of Linear Systems

Nonlinear Parametric Excitation

Applications

Future Directions

Bibliography

## Glossary

**Coexistence** The special case when all the independent solutions of a linear,  $T$ -periodic ODE are  $T$ -periodic.

**Hill's eq.** A second order ODE of the form  $\ddot{x} + p(t)x = 0$ , with  $p(t)$   $T$ -periodic.

**Instability pockets** Finite domains, usually intersections of instability tongues, where the trivial solution of linear,  $T$ -periodic ODEs is unstable.

**Instability tongues** Domains in parameter space where the trivial solution of linear,  $T$ -periodic ODEs is unstable.

**Mathieu equation** An ODE of the form  $\ddot{x} + (a + b \cos(t))x = 0$ .

**Parametric resonance** Resonance excitation arising for special values of coefficients, frequencies and other parameters in  $T$ -periodic ODEs.

**Quasi-periodic** A function of the form  $\sum_{i=1}^n f_i(t)$  with  $f_i(t)$   $T_i$ -periodic,  $n$  finite, and the periods  $T_i$  independent over  $\mathbb{R}$ .

**Sum resonance** A parametric resonance arising in the case of at least three frequencies in a  $T$ -periodic ODE.

## Definition of the Subject

Parametric resonance arises in mechanics in systems with external sources of energy and for certain parameter values. Typical examples are the pendulum with oscillating support and a more specific linearization of this pendulum, the Mathieu equation in the form

$$\ddot{x} + (a + b \cos(t))x = 0.$$

The time-dependent term represents the excitation. Tradition has it that parametric resonance is usually not considered in the context of systems with *external* excitation of the form  $\dot{x} = f(x) + \phi(t)$ , but for systems where time-dependence arises in the coefficients of the equation. Mechanically this means usually periodically varying stiffness, mass or load, in fluid or plasma mechanics one can think of frequency modulation or density fluctuation, in mathematical biology of periodic environmental changes. The term ‘parametric’ refers to the dependence on parameters and certain resonances arising for special values of the parameters. In the case of the Mathieu equation, the parameters are the frequency  $\omega$  ( $a = \omega^2$ ) of the equation without time-dependence and the excitation amplitude  $b$ ; see section “[Parametric Excitation of Linear Systems](#)” for an explicit demonstration of resonance phenomena in this two parameters system.

Mathematically the subject is concerned with ODEs with periodic coefficients. The study of linear dynamics of this type gave rise to a large amount of literature in the first half of the twentieth century and this highly technical, classical material is still accessible in textbooks. The standard equations are Hill’s equation and the Mathieu equation (see subsection “[Elementary](#)



Theory”). We will summarize a number of basic aspects. The reader is also referred to the article “Dynamics of Parametric Excitation” by Alan Champneys in this Encyclopedia.

Recently, the interest in nonlinear dynamics, new applications and the need to explore higher dimensional problems has revived the subject. Also structural stability and persistence problems have been investigated. Such problems arise as follows. Suppose that we have found a number of interesting phenomena for a certain equation and suppose we embed this equation in a family of equations by adding parameters or perturbations. Do the ‘interesting phenomena’ persist in the family of equations? If not, we will call the original equation structurally unstable. A simple example of structural instability is the harmonic equation which shows qualitative different behavior on adding damping. In general, Hamiltonian systems are structurally unstable in the wider context of dissipative dynamical systems.

## Introduction

Parametric resonance produces interesting mathematical challenges and plays an important part in many applications. The linear dynamics is already nontrivial whereas the nonlinear dynamics of such systems is extremely rich and largely unexplored. The role of symmetries is essential, both in linear and in nonlinear analysis. A classical example of parametric excitation is the swinging pendulum with oscillating support. The equation of motion describing the model is

$$\ddot{x} + (\omega_0^2 + p(t)) \sin x = 0, \quad (1)$$

where  $p(t)$  is a periodic function. Upon linearization – replacing  $\sin x$  by  $x$  – we obtain Hill’s equation (subsection “Elementary Theory”):

$$\ddot{x} + (\omega_0^2 + p(t))x = 0.$$

This equation was formulated around 1900 in the perturbation theory of periodic solutions in celestial mechanics. If we choose  $p(t) = \cos \omega t$ , Hill’s equation becomes the Mathieu equation. It

is well-known that special tuning of the frequency  $\omega_0$  and the period of excitation (of  $p(t)$ ) produces interesting instability phenomena (resonance). More generally we may study nonlinear parametric equations of the form

$$\ddot{x} + k\dot{x} + (\omega_0^2 + p(t))F(x) = 0, \quad (2)$$

where  $k > 0$  is the damping coefficient,  $F(x) = x + bx^2 + cx^3 + \dots$  and time is scaled so that  $p(t)$  is a  $p$ -periodic function with zero average. We may also take for  $p(t)$  a quasi-periodic or almost-periodic function. The books (Yakubovich and Starzhinskii 1975) cover most of the classical theory, but for a nice introduction see (Stoker 1950). In (Seyranian and Mailybaev 2003), emphasis is placed on the part played by parameters, it contains a rich survey of bifurcations of eigenvalues and various applications. There are many open questions for Eqs. (1) and (2); we shall discuss aspects of the classical theory, recent theoretical results and a few applications.

As noted before, in parametric excitation we have an oscillator with an independent source of energy. In examples, the oscillator is often described by a one degree of freedom system but of course many more degrees of freedom may play a part; see for instance in section “Applications” the case of coupled Mathieu-equations as studied in (Ruijgrok et al. 1993). In what follows,  $\varepsilon$  will always be a small, positive parameter.

## Perturbation Techniques

In this section we review the basic techniques to handle parametric perturbation problems. In the case of Poincaré–Lindstedt series which apply to periodic solutions, the expansions are in integer powers of  $\varepsilon$ . It should be noted that in general, other order functions of  $\varepsilon$  may play a part; see subsection “Elementary Theory” and (Verhulst 2005b).

### Poincaré–Lindstedt Series

One of the oldest techniques is to approximate a periodic solution by the construction of a convergent series in terms of the small parameter  $\varepsilon$ . The method can be used for equations of the form

$$\dot{x} = f(t, x) + \varepsilon g(t, x) + \varepsilon^2 \cdots,$$

with  $x \in \mathbb{R}^n$  and (usually) assuming that the ‘unperturbed’ problem  $\dot{y} = f(t, y)$  is understood and can be solved. Note that the method can also be applied to perturbed maps and difference equations. Suppose that the unperturbed problem contains a periodic solution, under what conditions can this solution be continued for  $\varepsilon > 0$ ? The answer is given by the conditions set by the implicit function theorem, see for formulations and theorems (Roseau 1966; Verhulst 1996). Usually we can associate with our perturbation problem a parameter space and one of the questions is then to find the domains of stability and instability. The common boundary of these domains is often characterized by the existence of periodic solutions and this is where Poincaré-Lindstedt series are useful. We will demonstrate this in the next section.

### Averaging

Averaging is a normalization method. In general, the term “normalization” is used whenever an expression or quantity is put in a simpler, standardized form. For instance, a  $n \times n$ -matrix with constant coefficients can be put in Jordan normal form by a suitable transformation. When the eigenvalues are distinct, this is a diagonal matrix.

Introductions to normalization can be found in (Arnold 1983; Golubitsky and Schaeffer 1985; Kuznetsov 2004) and (Cicogna and Gaeta 1999). For the relation between averaging and normalization in general the reader is referred to (Sanders et al. 2007; Verhulst 1996). For averaging in the so-called standard form it is assumed that we can put the perturbation problem in the form

$$\dot{x} = \varepsilon F(t, x) + \varepsilon^2 \cdots,$$

and that we have the existence of the limit

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T F(t, x) dt = F^0(x).$$

The analysis of the averaged equation  $\dot{y} = F^0(y)$  produces asymptotic approximations of

the solutions of the original equation on a long timescale; see (Sanders et al. 2007). Also, under certain conditions, critical points of the averaged equation correspond with periodic solutions in the original system. The choice to use Poincaré-Lindstedt series or the averaging method is determined by the amount of information one wishes to obtain. To find the location of stability and instability domains (the boundaries), Poincaré-Lindstedt series are very efficient. On the other hand, with somewhat more efforts, the averaging method will also supply this information with in addition the behavior of the solutions within the domains. For an illustration see subsection “Elementary Theory”.

### Resonance

Assume that  $x = 0$  is a critical point of the differential equation and write the system as:

$$\dot{x} = Ax + f(t, x, \varepsilon), \quad (3)$$

with  $x \in \mathbb{R}^n$ ,  $A$  a constant  $n \times n$ -matrix;  $f(t, x, \varepsilon)$  can be expanded in a Taylor series with respect to  $\varepsilon$  and in homogeneous vector polynomials in  $x$  starting with quadratic terms. Normalization of Eq. (3) means that by successive transformation we remove as many terms of Eq. (3) as possible. It would be ideal if we could remove all the nonlinear terms, i.e. linearize Eq. (3) by transformation. In general, however, some nonlinearities will be left and this is where resonance comes in. The eigenvalues  $\lambda_1, \dots, \lambda_n$  of the matrix  $A$  are resonant if for some  $i \in \{1, 2, \dots, n\}$  one has:

$$\sum_{j=1}^n m_j \lambda_j = \lambda_i, \quad (4)$$

with  $m_j \geq 0$  integers and  $m_1 + m_2 + \dots + m_n \geq 2$ . If the eigenvalues of  $A$  are non-resonant, we can remove all the nonlinear terms and so linearize the system. However, this is less useful than it appears, as in general the sequence of successive transformations to carry out the normalization will be divergent. The usefulness of normalization lies in removing nonresonant terms to a certain degree to simplify the analysis.

### Normalization of Time-Dependent Vectorfields

In problems involving parametric resonance, we have time-dependent systems such as equations perturbing the Mathieu equation. Details of proofs and methods to compute the normal form coefficients in such cases can be found in (Arnold 1983; Iooss and Adelmeyer 1992) and (Sanders et al. 2007). We summarize some aspects. Consider the following parameter and time dependent equation:

$$\dot{x} = F(x, \mu, t), \quad (5)$$

with  $x \in \mathbb{R}^m$  and the parameters  $\mu \in \mathbb{R}^p$ .

Here  $F(x, \mu, t) : \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R}^m$  is  $C^\infty$  in  $x$  and  $\mu$  and  $T$ -periodic in the  $t$ -variable. We assume that  $x = 0$  is a solution, so  $F(0, \mu, t) = 0$  and, moreover assume that the linear part of the vectorfield  $D_x F(0, 0, t)$  is time-independent for all  $t \in \mathbb{R}$ . We will write  $L_0 = D_x F(0, 0, t)$ . Expanding  $F(x, \mu, t)$  in a Taylor series with respect to  $x$  and  $\mu$  yields the equation:

$$\dot{x} = L_0 x + \sum_{n=2}^k F_n(x, \mu, t) + O(|(x, \mu)|^{k+1}), \quad (6)$$

where the  $F_n(x, \mu, t)$  are homogeneous polynomials in  $x$  and  $\mu$  of degree  $n$  with  $T$ -periodic coefficients.

**Theorem 1** *Let  $k \in \mathbb{N}$ . There exists a (parameter- and time-dependent) transformation*

$$x = \hat{x} + \sum_{n=2}^k P_n(\hat{x}, \mu, t),$$

where  $P_n(\hat{x}, \mu, t)$  are homogeneous polynomials in  $x$  and  $\mu$  of degree  $n$  with  $T$ -periodic coefficients, such that Eq. (6) takes the form (dropping the hat):

$$\begin{aligned} \dot{x} &= L_0 x + \sum_{n=2}^k \tilde{F}_n(x, \mu, t) + O(|(x, \mu)|^{k+1}) \\ \dot{\mu} &= 0. \end{aligned} \quad (7)$$

The truncated vectorfield:

$$\dot{x} = L_0 x + \sum_{n=2}^k \tilde{F}_n(x, \mu, t) = \tilde{F}(x, \mu, t), \quad (8)$$

which will be called the normal form of Eq. (5), has the following properties:

1.  $\frac{d}{dt} e^{L_0 t} \tilde{F}(e^{-L_0 t} x, \mu, t) = 0$ , for all  $(x, \mu) \in \mathbb{R}^{m+p}$ ,  $t \in \mathbb{R}$ .
2. If Eq. (5) is invariant under an involution (i. e.  $SF(x, \mu, t) = F(Sx, \mu, t)$  with  $S$  an invertible linear operator such that  $S^2 = I$ ), then the truncated normal form (8) is also invariant under  $S$ . Similarly, if Eq. (5) is reversible under an involution  $R$  (i.e.  $RF(x, \mu, t) = -F(Rx, \mu, t)$ ), then the truncated normal form (8) is also reversible under  $R$ .

For a proof, see (Iooss and Adelmeyer 1992).

The theorem will be applied to situations where  $L_0$  is semi-simple and has only purely imaginary eigenvalues. We take  $L_0 = \text{diag} \{i\lambda_1, \dots, i\lambda_m\}$ . In our applications,  $m = 2l$  is even and  $\lambda_{l+j} = -\lambda_j$  for  $j = 1, \dots, l$ . The variable  $x$  is then often written as  $x = (z_1, \dots, z_l, \bar{z}_1, \dots, \bar{z}_l)$ .

Assume  $L_0 = \text{diag} \{i\lambda_1, \dots, i\lambda_m\}$  then:

- A term  $x_1^{\gamma_1} \dots x_m^{\gamma_m} e^{i\frac{2\pi}{T}kt}$  is in the  $j$ th component of the Taylor–Fourier series of  $\tilde{F}(x, \mu, t)$  if:

$$-\lambda_j + \frac{2\pi}{T}k + \gamma_1 \lambda_1 + \dots + \gamma_m \lambda_m = 0. \quad (9)$$

This is known as the resonance condition.

- Transforming the normal form through  $x = e^{L_0 t} w$  leads to an autonomous equation for  $w$ :

$$\dot{w} = \sum_{n=2}^k \tilde{F}_n(w, \mu, 0). \quad (10)$$

- An important result is this: If Eq. (5) is invariant (respectively reversible) under an involution  $S$ , then this also holds for Eq. (10).
- The autonomous normal form (10) is invariant under the action of the group  $\mathcal{G} = \{g | gx = e^{jL_0 T} x, j \in \mathbb{Z}\}$ , generated by  $e^{L_0 T}$ .

Note that this group is discrete if the ratios of the  $\lambda_i$  are rational and continuous otherwise.

For a proof of the last two statements see (Ruijgrok 1995).

By this procedure we can make the system autonomous. This is very effective as the autonomous normal form (10) can be used to prove the existence of periodic solutions and invariant tori of Eq. (5) near  $x = 0$ . We have:

**Theorem 2** *Let  $\varepsilon > 0$ , sufficiently small, be given. Scale  $w = \varepsilon \hat{w}$ .*

1. *If  $\hat{w}_0$  is a hyperbolic fixed point of the (scaled) Eq. (10), then Eq. (5) has a hyperbolic periodic solution  $x(t) = \varepsilon \hat{w}_0 + O(\varepsilon^{k+1})$ .*
2. *If the scaled Eq. (10) has a hyperbolic closed orbit, then Eq. (5) has a hyperbolic invariant torus.*

These results are related to earlier theorems in (Bogoliubov and Mitropolskii 1961), see also the survey (Verhulst 2005a). Later we shall discuss normalization in the context of the so-called sum-resonance.

### Remarks on Limit Sets

In studying a dynamical system the behavior of the solutions is for a large part determined by the limit sets of the system. The classical limit sets are equilibria and periodic orbits. Even when restricting to autonomous equations of dimension three, we have no complete classification of possible limit sets and this makes the recognition and description of non-classical limit sets important. In parametrically excited systems, the following limit sets, apart from the classical ones, are of interest:

- Chaotic attractors. Various scenarios were found, see (Ruijgrok 1995; Ruijgrok and Verhulst 1996; Tondl et al. 2000).
- Strange attractors without chaos, see (Pikovsky and Feudel 1995). The natural presence of various forcing periods in real-life models make their occurrence quite plausible.
- Attracting tori. These limit sets are not difficult to find; they arise for instance as a consequence

of a Neimark–Sacker bifurcation of a periodic solution, see (Kuznetsov 2004).

- Attracting heteroclinic cycles, see (Krupa 1997).

A large number of these phenomena can be studied both by numerics *and* by perturbation theory; using the methods simultaneously gives additional insight.

### Parametric Excitation of Linear Systems

As we have seen in the introduction, parametric excitation leads to the study of second order equations with periodic coefficients. More in general such equations arise from linearization near  $T$ -periodic solutions of  $T$ -periodic equations of the form  $\dot{y} = f(t, y)$ . Suppose  $y = \phi(t)$  is a  $T$ -periodic solution; putting  $y = \phi(t) + x$  produces upon linearization the  $T$ -periodic equation

$$\dot{x} = f_x(t, \phi(t))x. \quad (11)$$

This equation often takes the form

$$\dot{x} = Ax + \varepsilon B(t)x, \quad (12)$$

in which  $x \in \mathbb{R}^m$ ;  $A$  is a constant  $m \times m$ -matrix,  $B(t)$  is a continuous,  $T$ -periodic  $m \times m$ -matrix,  $\varepsilon$  is a small parameter. For elementary studies of such an equation, the Poincaré–Lindstedt method or continuation method is quite efficient. The method applies to nonlinear equations of arbitrary dimension, but we shall demonstrate its use for equations of Mathieu type.

### Elementary Theory

Floquet theory tells us that the solutions of Eq. (12) can be written as:

$$x(t) = \Phi(t, \varepsilon)e^{C(\varepsilon)t}, \quad (13)$$

with  $\Phi(t, \varepsilon)$  a  $T$ -periodic  $m \times m$ -matrix,  $C(\varepsilon)$  a constant  $m \times m$ -matrix and both matrices having an expansion in order functions of  $\varepsilon$ . The determination of  $C(\varepsilon)$  provides us with the stability behavior of the solutions. A particular case of Eq. (12) is Hill's equation:

$$\ddot{x} + b(t, \varepsilon)x = 0, \quad (14)$$

which is of second order;  $b(t, \varepsilon)$  is a scalar  $T$ -periodic function. A number of cases of Hill's equation are studied in (Magnus and Winkler 1966). A particular case of Eq. (14) which arises frequently in applications is the Mathieu equation:

$$\ddot{x} + (\omega^2 + \varepsilon \cos 2t)x = 0, \omega > 0, \quad (15)$$

which is reversible. (In (Magnus and Winkler 1966) one also finds Lamé's, Ince's, Hermite's, Whittaker–Hill and other Hill equations.) A typical question is: for which values of  $\omega$  and  $\varepsilon$  in  $(\omega^2, \varepsilon)$ -parameter space is the trivial solution  $x = \dot{x} = 0$  stable?

Solutions of Eq. (15) can be written in the Floquet form (13), where in this case  $\Phi(t, \varepsilon)$  will be  $p$ -periodic. The eigenvalues  $\lambda_1, \lambda_2$  of  $C$ , which are called characteristic exponents and are  $\mathcal{E}$ -dependent, determine the stability of the trivial solution. For the characteristic exponents of Eq. (12) we have:

$$\sum_{i=1}^n \lambda_i = \frac{1}{T} \int_0^T \text{Tr}(A + \varepsilon B(t)) dt, \quad (16)$$

see Theorem 6.6 in (Verhulst 1996). So in the case of Eq. (15) we have:

$$\lambda_1 + \lambda_2 = 0. \quad (17)$$

The exponents are functions of  $\mathcal{E}$ ,  $\lambda_1 = \lambda_1(\varepsilon)$ ,  $\lambda_2 = \lambda_2(\varepsilon)$  and clearly  $\lambda_1(0) = i\omega$ ,  $\lambda_2(0) = -i\omega$ . As  $\lambda_1(\varepsilon) = -\lambda_2(\varepsilon)$ , the characteristic exponents, which are complex conjugate, are purely imaginary or real. The implication is that if  $\omega^2 \neq n^2$ ,  $n = 1, 2, \dots$  the characteristic exponents are purely imaginary and  $x = 0$  is stable near  $\varepsilon = 0$ . If  $\omega^2 = n^2$  for some  $n \in \mathbb{N}$ , however, the imaginary part of  $\exp(C(\varepsilon)t)$  can be absorbed into  $\Phi(t, \varepsilon)$  and the characteristic exponents may be real. We assume now that  $\omega^2 = n^2$  for some  $n \in \mathbb{N}$ , or near this value, and we shall look for periodic solutions of  $x(t)$  of Eq. (15) as these solutions define the boundaries between stable and unstable solutions. We put:

$$\omega^2 = n^2 - \varepsilon\beta, \quad (18)$$

with  $\beta$  a constant, and we apply the Poincaré–Lindstedt method to find the periodic solutions; see Appendix 2 in (Verhulst 1996). We find that periodic solutions exist for  $n = 1$  if:

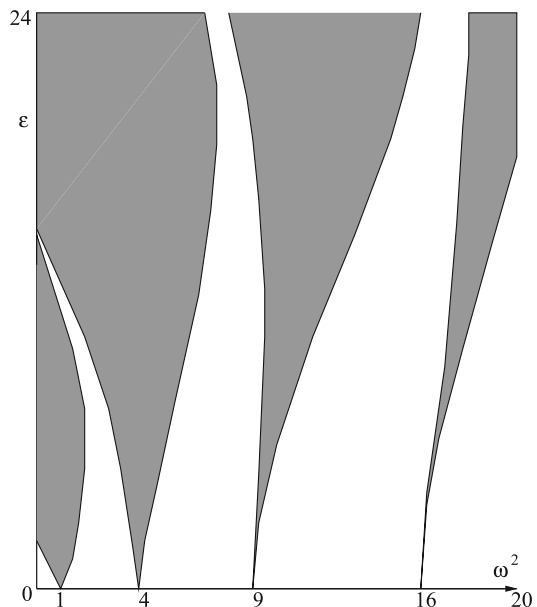
$$\omega^2 = 1 \pm \frac{1}{2}\varepsilon + O(\varepsilon^2).$$

In the case  $n = 2$ , periodic solutions exist if:

$$\begin{aligned} \omega^2 &= 4 - \frac{1}{48}\varepsilon^2 + O(\varepsilon^4), \\ \omega^2 &= 4 + \frac{5}{48}\varepsilon^2 + O(\varepsilon^4). \end{aligned} \quad (19)$$

The corresponding instability domains are called Floquet tongues, instability tongues or resonance tongues, see Fig. 1.

On considering higher values of  $n$ , we have to calculate to a higher order of  $\varepsilon$ . At  $n = 1$  the boundary curves are intersecting at positive angles at  $\varepsilon = 0$ , at  $n = 2$  ( $\omega^2 = 4$ ) they are tangent; the order of tangency increases as  $n - 1$  (contact



**Perturbation Analysis of Parametric Resonance, Fig. 1** Floquet tongues of the Mathieu Eq. (15); the instability domains are shaded

of order  $n$ ), making instability domains more and more narrow with increasing resonance number  $n$ .

### Higher Order Approximation and an Unexpected Timescale

The instability tongue of the Mathieu equation at  $n = 1$  can be determined with more precision by Poincaré expansion. On using averaging, one also characterizes the flow outside the tongue boundary and this results in a surprise. Consider Eq. (15) in the form

$$\ddot{x} + (1 + \varepsilon a + \varepsilon^2 b + \varepsilon \cos 2t)x = 0,$$

where we can choose  $a = \pm \frac{1}{2}$  to put the frequency with first order precision at the tongue boundary. The eigenvalues of the trivial solution are from first order averaging

$$\lambda_{1,2} = \pm \frac{1}{2} \sqrt{\frac{1}{4} - a^2},$$

which agrees with Poincaré expansion;  $a^2 > \frac{1}{4}$  gives stability, the  $<$  inequality instability. The transition value  $a^2 = \frac{1}{4}$  gives the tongue location. Take for instance the  $+$  sign. Second order averaging, see (Verhulst 2005b), produces for the eigenvalues of the trivial solution

$$\lambda_{1,2} = \pm \sqrt{-\frac{1}{4} \left( \frac{1}{32} + b \right) \varepsilon^3 + \left( \frac{1}{64} + \frac{1}{2} b \right) \cdot \left( \frac{7}{64} - \frac{1}{2} b \right) \varepsilon^4}.$$

So, if  $\frac{1}{32} + b > 0$  we have stability, if  $\frac{1}{32} + b < 0$  instability; at  $b = -\frac{1}{32}$  we have the second order approximation of this tongue boundary. Note that near this boundary the solutions are characterized by eigenvalues of  $O(\varepsilon^{\frac{3}{2}})$  and accordingly the time-dependence by timescale  $\varepsilon^{\frac{3}{2}}t$ .

### The Mathieu Equation with Viscous Damping

In real-life applications there is always the presence of damping. We shall consider the effect of its simplest form, small viscous damping. Eq. (15) is extended by adding a linear damping term:

$$\ddot{x} + \kappa \dot{x} + (\omega^2 + \varepsilon \cos 2t)x = 0, \quad a, \kappa > 0. \quad (20)$$

We assume that the damping coefficient is small,  $\kappa = \varepsilon \kappa_0$ , and again we put  $\omega^2 = n^2 - \varepsilon \beta$  to apply the Poincaré-Lindstedt method.

We find periodic solutions in the case  $n = 1$  if:

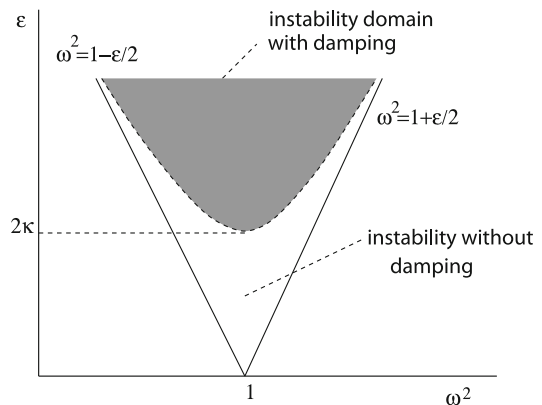
$$\omega^2 = 1 \pm \sqrt{\frac{1}{4} \varepsilon^2 - \kappa^2}. \quad (21)$$

Relation (21) corresponds with the curve of periodic solutions, which in  $(\omega^2, \varepsilon)$ -parameter space separates stable and unstable solutions. We observe the following phenomena. If  $0 < \kappa < \frac{1}{2} \varepsilon$ , we have an instability domain which by damping has been lifted from the  $\omega^2$ -axis; also *the width has shrunk*. If  $\kappa > \frac{1}{2} \varepsilon$  the instability domain has vanished. For an illustration see Fig. 2.

Repeating the calculations for  $n \geq 2$  we find no instability domains at all; damping of  $O(\varepsilon)$  stabilizes the system for  $\varepsilon$  small. To find an instability domain we have to decrease the damping, for instance if  $n = 2$  we have to take  $\kappa = \varepsilon^2 \kappa_0$ .

### Coexistence

Linear periodic equations of the form (12) have  $m$  independent solutions and it is possible that all the independent solutions are periodic. This is called ‘coexistence’ and one of the consequences



**Perturbation Analysis of Parametric Resonance, Fig. 2** First order approximation of instability domains without and with damping for Eq. (20) near  $\omega^2 = 1$



is that the instability tongues vanish. An example is Ince's equation:

$$(1 + a \cos t)\ddot{x} + \kappa \sin t \dot{x} + (\omega^2 + \varepsilon \cos t)x = 0,$$

see (Magnus and Winkler 1966). An interesting question is whether this phenomenon persists under nonlinear perturbations; we return to this question in subsection “Coexistence Under Nonlinear Perturbation”.

### More General Classical Results

The picture presented by the Mathieu equation resulting in resonance tongues in the  $\omega$ ,  $\varepsilon$ -parameter space, stability and instability intervals as parametrized by  $\omega$  shown in Fig. 1, has been studied for more general types of Hill's equation. The older literature can be found in (Strutt 1932), see also (Van der Pol and Strutt 1928).

Consider Hill's equation in the form

$$\ddot{x} + (\omega^2 + \varepsilon f(t))x = 0, \quad (22)$$

with  $f(t)$  periodic and represented by a Fourier series. Along the  $\omega^2$ -axis there exist instability intervals of size  $L_m$ , where  $m$  indicates the  $m$ th instability interval. In the case of the Mathieu equation, we have from (Hale 1963)

$$L_m = O(\varepsilon^m).$$

The resonance tongues become increasingly narrow.

For general periodic  $f(t)$  we have weak estimates, like  $L_m = O(\varepsilon)$ , but if we assume that the Fourier series is finite, the estimates can be improved. Put

$$f(t) = \sum_{j=0}^s f_j \cos 2jt,$$

so  $f(t)$  is even and  $p$ -periodic. From (Levy and Keller 1963) we have the following estimates:

- If we can write  $m = sp$  with  $p \in \mathbb{N}$ , we have

$$L_m = \frac{8s^2}{((p-1)!)^2} \left( \frac{|f_s \varepsilon|}{8s^2} \right)^p + O(\varepsilon^{p+1}).$$

- If we can not decompose  $m$  like this and  $sp < m < s(p+1)$ , we have

$$L_m = O(\varepsilon^{p+1}).$$

In the case of Eq. (22) we have no dissipation and then it can be useful to introduce canonical transformations and Poincaré maps. In this case, for example, put

$$\dot{x} = y, \dot{y} = -\frac{\partial H}{\partial x},$$

with Hamiltonian function

$$H(x, y, t) = \frac{1}{2}y^2 + \frac{1}{2}(\omega^2 + \varepsilon f(t))x^2.$$

We can split  $H = H_0 + \varepsilon H_1$  with  $H_0 = \frac{1}{2} \times (y^2 + \omega^2 x^2)$  and apply canonical perturbation theory. Examples of this line of research can be found in (Broer and Levi 1995; Broer and Simó 2000). Interesting conclusions can be drawn with respect to the geometry of the resonance tongues, crossings of tongues and as a possible consequence the presence of so-called instability pockets. In this context, the classical Mathieu equation turns out to be quite degenerate.

Hill's equation in the case of damping was considered in (Seyranian 2001); see also (Seyranian and Mailybaev 2003) where an arbitrary number of degrees of freedom is discussed.

### Quasi-Periodic Excitation

Equations of the form

$$\ddot{x} + (\omega^2 + \varepsilon p(t))x = 0, \quad (23)$$

with parametric excitation quasi-periodic or almost-periodic, arise often in applications. Floquet theory does not apply in this case but we can still use perturbation theory. A typical example would be two rationally independent frequencies:



$$p(t) = \cos t + \cos \gamma t,$$

with  $\gamma$  irrational. As an interesting example, in (Broer and Simó 1998),  $\gamma = \frac{1}{2}(1 + \sqrt{5})$  was chosen, the golden number. It will be no surprise that many more complications arise for large values of  $\epsilon$ , but for  $\epsilon$  small (the assumption in this article), the analysis runs along similar lines producing resonance tongues, crossings of tongues and instability pockets. See also extensions in (Broer et al. 2003). Detailed perturbation expansions are presented in (Zounes and Rand 1998) with a comparison of Poincaré expansion, the harmonic balance method and numerics; there is good agreement between the methods. Real-life models contain dissipation which inspired the authors of (Zounes and Rand 1998) to consider the equation

$$\ddot{x} + 2\mu\dot{x} + (\omega^2 + \epsilon(\cos t + \cos \gamma t))x = 0, \\ \mu > 0,$$

$\gamma$  irrational. They conclude that

- The instability tongues become thinner and recede into the  $\mathcal{E}$ -axis as  $\mu$  increases.
- High-order resonance tongues seem to be more affected by dissipation than low-order ones producing a dramatic loss of ‘fine detail’, even for small  $\mu$ .
- The results of varying the parameter  $\mu$  certainly needs more investigation.

### Parametrically Forced Oscillators in Sum Resonance

In applications where more than one degree of freedom plays a part, many more resonances are possible. For a number of interesting cases and additional literature see (Seyranian and Mailybaev 2003). An important case is the so-called sum resonance. In (Hoveijn and Ruijgrok 1995) a geometrical explanation is presented for the phenomena in this case using ‘all’ the parameters as unfolding parameters. It will turn out that four parameters are needed to give a complete description. Fortunately three suffice to visualize the situation. Consider the following type of differential equation with three frequencies

$$\dot{z} = Az + \epsilon f(z, \omega_0 t; \lambda), \quad z \in \mathbb{R}^4, \quad \lambda \in \mathbb{R}^p, \quad (24)$$

which describes a system of two parametrically forced coupled oscillators. Here  $A$  is a  $4 \times 4$  matrix, containing parameters, and with purely imaginary eigenvalues  $\pm i\omega_1$  and  $\pm i\omega_2$ . The vector valued function  $f$  is  $2\pi$ -periodic in  $\omega_0 t$  and  $f(0, \omega_0 t; \lambda) = 0$  for all  $t$  and  $\lambda$ . Eq. (24) can be resonant in many different ways. We consider the sum resonance

$$\omega_1 + \omega_2 = \omega_0,$$

where the system may exhibit instability. The parameter  $\lambda$  is used to control detuning  $\delta = (\delta_1, \delta_2)$  of the frequencies  $(\omega_1, \omega_2)$  from resonance and damping  $\mu = (\mu_1, \mu_2)$ . We summarize the analysis from (Hoveijn and Ruijgrok 1995).

- The first step is to put Eq. (24) into normal form by normalization or averaging. In the normalized equation the time-dependence appears only in the higher order terms. But the autonomous part of this equation contains enough information to determine the stability regions of the origin. The linear part of the normal form is  $\dot{z} = A(\delta, \mu)z$  with

$$A(\delta, \mu) = \begin{pmatrix} B(\delta, \mu) & 0 \\ 0 & \overline{B}(\delta, \mu) \end{pmatrix}, \quad (25)$$

and

$$B(\delta, \mu) = \begin{pmatrix} i\delta_1 - \mu_1 & \alpha_1 \\ \overline{\alpha}_2 & -i\delta_2 - \mu_2 \end{pmatrix}. \quad (26)$$

Since  $A(\delta, \mu)$  is the complexification of a real matrix, it commutes with complex conjugation. Furthermore, according to the normal form Theorem 1 and if  $\omega_1$  and  $\omega_2$  are independent over the integers, the normal form of Eq. (24) has a continuous symmetry group.

- The second step is to test the linear part  $A(\delta, \mu)$  of the normalized equation for structural stability i.e. to answer the question whether there

exist open sets in parameter space where the dynamics is qualitatively the same. The family of matrices  $A(\delta, \mu)$  is parametrized by the detuning  $\delta$  and the damping  $\mu$ . We first identify the most degenerate member  $N$  of this family and then show that  $A(\delta, \mu)$  is its versal unfolding in the sense of (Arnold 1983). The family  $A(\delta, \mu)$  is equivalent to a versal unfolding  $U(\lambda)$  of the degenerate member  $N$ .

- Put differently, the family  $A(\delta, \mu)$  is structurally stable for  $\delta, \mu > 0$ , whereas  $A(\delta, 0)$  is not. This has interesting consequences in applications as small damping and zero damping may exhibit very different behavior, see section “Rotor Dynamics”. In parameter space, the stability regions of the trivial solution are separated by a *critical surface* which is the hypersurface where  $A(\delta, \mu)$  has at least one pair of purely imaginary complex conjugate eigenvalues. This critical surface is diffeomorphic to the *Whitney umbrella*, see Fig. 3 and for references (Hoveijn and Ruijgrok 1995). It is the singularity of the Whitney umbrella that causes the discontinuous behavior of the stability diagram in section “Rotor Dynamics”. The structural

stability argument guarantees that the results are ‘universally valid’, i.e. they qualitatively hold for generic systems in sum resonance.

## Nonlinear Parametric Excitation

Adding nonlinear effects to parametric excitation strongly complicates the dynamics. We start with adding nonlinear terms to the (generalized) Mathieu equation. Consider the following equation that includes dissipation:

$$\ddot{x} + \kappa \dot{x} + (\omega^2 + \varepsilon p(t))f(x) = 0, \quad (27)$$

where  $\kappa > 0$  is the damping coefficient,  $f(x) = x + bx^2 + cx^3 + \dots$ , and time is scaled so that:

$$p(t) = \sum_{l \in \mathbb{Z}} a_{2l} e^{2ilt}, \quad a_0 = 0, \quad a_{-2l} = \bar{a}_{2l}, \quad (28)$$

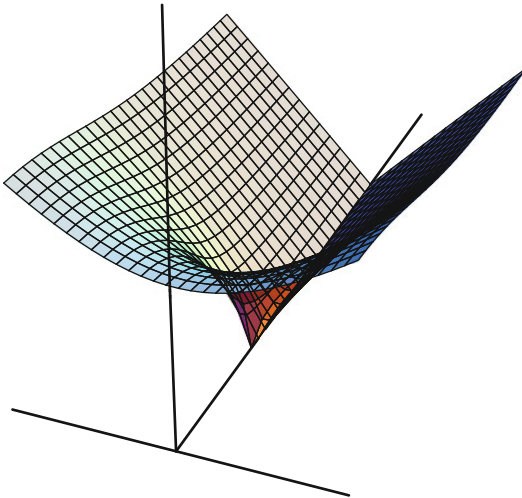
is an even  $\pi$ -periodic function with zero average. As we have seen in section “Parametric Excitation of Linear Systems”, the trivial solution  $x = 0$  is unstable when  $\kappa = 0$  and  $\omega^2 = n^2$ , for all  $n \in \mathbb{N}$ . Fix a specific  $n \in \mathbb{N}$  and assume that  $\omega_2$  is close to  $n^2$ . We will study the bifurcations from the solution  $x = 0$  in the case of primary resonance, which by definition occurs when the Fourier expansion of  $p(t)$  contains nonzero terms  $a_{2n} e^{2int}$  and  $a_{-2n} e^{-2int}$ . The bifurcation parameters in this problem are the detuning  $\sigma = \omega^2 - n^2$ , the damping coefficient  $\kappa$  and the Fourier coefficients of  $p(t)$ , in particular  $a_{2n}$ . The Fourier coefficients are assumed to be of equal order of magnitude.

### The Conservative Case, $\kappa = 0$

An early paper is (Len and Rand 1988) in which Eq. (27) for  $\kappa = 0$  is associated with the Hamiltonian

$$H(x, \dot{x}, t) = \frac{1}{2} \dot{x}^2 + \frac{\omega^2}{2} x^2 + p(t) \int_0^x f(s) ds.$$

After transformation of the Hamiltonian, Lie transforms are implemented by MACSYMA to produce normal form approximations to  $O(\varepsilon^2)$ . A number of examples show interesting bifurcations.



**Perturbation Analysis of Parametric Resonance, Fig. 3** The critical surface in  $(\mu_+, \mu_-, \delta_+)$  space for Eq. (24).  $\mu_+ = \mu_1 + \mu_2$ ,  $\mu_- = \mu_1 - \mu_2$ ,  $\delta_+ = \delta_1 + \delta_2$ . Only the part  $\mu_+ > 0$  and  $\delta_+ > 0$  is shown. The parameters  $\delta_1, \delta_2$  control the detuning of the frequencies, the parameters  $\mu_1, \mu_2$  the damping of the oscillators (vertical direction). The base of the umbrella lies along the  $\delta_+$ -axis

A related approach can be found in Broer and Vegter (1992); as  $p(t)$  is even, the equation is time-reversible. After construction of the Poincaré (timeperiodic) map, normal forms are obtained by equivariant transformations. This leads to a classification of integrable normal forms that are approximations of the family of Poincaré maps, a family as the map is parametrized by  $\omega$  and the coefficients of  $p(t)$ .

Interestingly, the nonlinearity  $\alpha x^3$  is combined with the quasi-periodic Mathieu equation in Zounes and Rand (2002) where global phenomena are described like resonance bands and chaos.

### Adding Dissipation, $\kappa > 0$

Again time-periodic normal form calculations are used to approximate the dynamics; see (Ruijgrok 1995), also (Ruijgrok and Verhulst 1996) and the monograph (Tondl et al. 2000). The reflection symmetry in the normal form equations implies that all fixed points come in pairs, and that bifurcations of the origin will be symmetric (such as pitchfork bifurcations). We observe that the normal form equations show additional symmetries if either  $f(x)$  in Eq. (27) is odd in  $x$  or if  $n$  is odd. The general normal form can be seen as a non-symmetric perturbation of the symmetric case. One finds pitchfork and saddle-node bifurcations, in fact all codimension one bifurcations; for details and pictures see Chap. 9 in Tondl et al. (2000).

### Coexistence Under Nonlinear Perturbation

A model describing free vibrations of an elastica is described in Ng and Rand (2003):

$$\left(1 - \frac{\varepsilon}{2} \cos 2t\right) \ddot{x} + \varepsilon \sin 2t \dot{x} + cx + \varepsilon \alpha x^2 = 0.$$

For  $\alpha = 0$ , the equation shows the phenomenon of co-existence. It is shown by second order averaging in (Ng and Rand 2003) that for  $\alpha \neq 0$  there exist open sets of parameter values for which the trivial solution is unstable.

An application to the stability problem of a family of periodic solutions in a Hamiltonian system is given in (Recktenwald and Rand 2005).

### Other Nonlinearities

In applications various nonlinear terms play a part. In (Ng and Rand 2002) one considers

$$\begin{aligned} \ddot{x} + (\omega^2 + \varepsilon \cos(t)) \\ + \varepsilon (Ax^3 + Bx^2\dot{x} + Cx\dot{x}^2 + D\dot{x}^3) \\ = 0, \end{aligned}$$

where averaging is applied near the 2 : 1-resonance. If  $B, D < 0$  the corresponding terms can be interpreted as progressive damping. It turns out that for a correct description of the bifurcations second-order averaging is needed.

Nonlinear damping can be of practical interest. The equation

$$\ddot{x} + (\omega^2 + \varepsilon \cos(t)) + \mu |\dot{x}| \dot{x} = 0,$$

is studied with  $\mu$  also a small parameter. A special feature is that an acceptable description of the phenomena can be obtained in a semi-analytical way by using Mathieu-functions as starting point. The analysis involves the use of Padé-approximants, see (Ramani et al. 2004).

## Applications

There are many applications of parametric resonance, in particular in engineering. In this section we consider a number of significant applications, but of course without any attempt at completeness. See also (Seyranian and Mailybaev 2003) and the references in the additional literature.

### The Parametrically Excited Pendulum

Choosing the pendulum case  $f(x) = \sin(x)$  in Eq. (27) we have

$$\ddot{x} + \kappa \dot{x} + (\omega^2 + \varepsilon p(t)) \sin(x) = 0.$$

It is natural, because of the  $\sin$  periodicity, to analyze the Poincaré map on the cylindrical section  $t = 0 \bmod 2\pi \mathbb{Z}$ . This map has both a spatial and a temporal symmetry. As we know from the preceding section, perturbation theory applied near the equilibria  $x = 0, x = \pi$ , produces

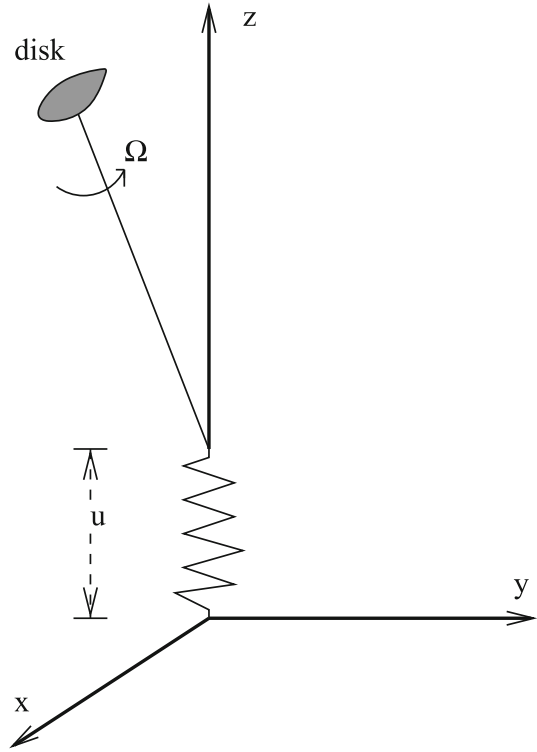
integrable normal forms. For larger excitation (larger values of  $\varepsilon$ ), the system exhibits the usual picture of Hamiltonian chaos; for details see (Broer et al. 2005; McLaughlin 1981).

The inverted case is intriguing. It is well-known that the upper equilibrium of an undamped pendulum can be stabilized by vertical oscillations of the suspension point with certain frequencies. See for references (Broer et al. 1998; Broer et al. 1999) and (Seyranian and Seyranian 2006). In (Broer et al. 1998) the genericity of the classical result is studied for (conservative) perturbations respecting the symmetries of the equation. In (Broer et al. 1999) genericity is studied for (conservative) perturbations where the spatial symmetry is broken, replacing  $\sin x$  by more general  $2\pi$ -periodic functions. Stabilization is still possible but the dynamics is more complicated.

### Rotor Dynamics

When adding linear damping to a system there can be a striking discontinuity in the bifurcational behavior. Phenomena like this have already been observed and described in for instance (Yakubovich and Starzhinskii 1975) or (Szemplinska-Stupnicka 1990). The discontinuity is a fundamental structural instability in linear gyroscopic systems with at least two degrees of freedom and with linear damping. The following example is based on (Ruijgrok et al. 1993; Tondl 1991a).

Consider a rigid rotor consisting of a heavy disk of mass  $M$  which is rotating with rotationspeed  $\Omega$  around an axis. The axis of rotation is elastically mounted on a foundation; the connections which are holding the rotor in an upright position are also elastic. To describe the position of the rotor we have the axial displacement  $u$  in the vertical direction (positive upwards), the angle of the axis of rotation *with respect to the*  $z$ -axis and *around* the  $z$ -axis. Instead of these two angles we will use the projection of the center of gravity motion on the horizontal  $(x, y)$ -plane, see Fig. 4. Assuming small oscillations in the upright ( $u$ ) position, frequency  $2\eta$ , the equations of motion become:



**Perturbation Analysis of Parametric Resonance, Fig. 4** Rotor with diskmass  $M$ , elastically mounted with axial ( $u$ ) and lateral directions

$$\begin{aligned}\ddot{x} + 2\alpha\dot{y} + (1 + 4\varepsilon\eta^2 \cos 2\eta t)x &= 0 \\ \ddot{y} - 2\alpha\dot{x} + (1 + 4\varepsilon\eta^2 \cos 2\eta t)y &= 0\end{aligned}\quad (29)$$

System (29) constitutes a system of Mathieu-like equations, where we have neglected the effects of damping. Abbreviating  $P(t) = 4\eta^2 \cos 2\eta t$ , the corresponding Hamiltonian is:

$$\begin{aligned}H &= \frac{1}{2}(1 + \alpha^2 + \varepsilon P(t))x^2 + \frac{1}{2}p_x^2 + \frac{1}{2} \\ &\quad \times (1 + \alpha^2 + \varepsilon P(t))y^2 + \frac{1}{2}p_y^2 + \alpha x p_y \\ &\quad - \alpha y p_x,\end{aligned}$$

where  $p_x, p_y$  are the momenta. The natural frequencies of the unperturbed system (29),  $\varepsilon = 0$ , are  $\omega_1 = \sqrt{\alpha^2 + 1} + \alpha$  and  $\omega_2 = \sqrt{\alpha^2 + 1} - \alpha$ . By putting  $z = x + iy$ , system (29) can be written as:

$$\ddot{z} - 2\alpha i \dot{z} + (1 + 4\epsilon\eta^2 \cos 2\eta t)z = 0. \quad (30)$$

Introducing the new variable:

$$v = e^{-i\alpha t} z, \quad (31)$$

and putting  $\eta t = \tau$ , we obtain:

$$v'' + \left( \frac{1 + \alpha^2}{\eta^2} + 4\epsilon \cos 2\tau \right) v = 0, \quad (32)$$

where the prime denotes differentiation with respect to  $\tau$ . By writing down the real and imaginary parts of this equation, we get two identical Mathieu equations. We conclude that the trivial solution is stable for  $\epsilon$  small enough, providing that  $\sqrt{1 + \alpha^2}$  is not close to  $n\eta$ , for some  $n = 1, 2, 3, \dots$ . The first-order interval of instability,  $n = 1$ , arises if:

$$\sqrt{1 + \alpha^2} \approx \eta. \quad (33)$$

If condition (33) is satisfied, the trivial solution of Eq. (32) is unstable. Therefore, the trivial solution of system (29) is also unstable. Note that this instability arises when:

$$\omega_1 + \omega_2 = 2\eta,$$

i.e. when the sum of the eigenfrequencies of the unperturbed system equals the excitation frequency  $2\eta$ . This is known as a sum resonance of first order. The domain of instability can be calculated as in subsection “[Elementary Theory](#)”; we find for the boundaries:

$$\eta_b = \sqrt{1 + \alpha^2} (1 \pm \epsilon) + O(\epsilon^2). \quad (34)$$

The second order interval of instability of Eq. (32),  $n = 2$ , arises when:

$$\sqrt{1 + \alpha^2} \approx 2\eta, \quad (35)$$

i.e.  $\omega_1 + \omega_2 \approx \eta$ . This is known as a sum resonance of second order. As above, we find the boundaries of the domains of instability:

$$\begin{aligned} 2\eta &= \sqrt{1 + \alpha^2} \left( 1 + \frac{1}{24} \epsilon^2 \right) + O(\epsilon^4), \\ 2\eta &= \sqrt{1 + \alpha^2} \left( 1 - \frac{5}{24} \epsilon^2 \right) + O(\epsilon^4). \end{aligned} \quad (36)$$

Higher order combination resonances can be studied in the same way; the domains of instability in parameter space continue to narrow as  $n$  increases. It should be noted that the parameter  $\alpha$  is proportional to the rotation speed  $\Omega$  of the disk and to the ratio of the moments of inertia.

### Instability by Damping

We add small linear damping to system (29), with positive damping parameter  $\mu = 2\epsilon\kappa$ . This leads to the equation:

$$\ddot{z} - 2\alpha i \dot{z} + (1 + 4\epsilon\eta^2 \cos 2\eta t)z + 2\epsilon\kappa \dot{z} = 0. \quad (37)$$

Because of the damping term, we can no longer reduce the complex Eq. (37) to two identical second order real equations, as we did in the previous section. In the sum resonance of the first order, we have  $\omega_1 + \omega_2 \approx 2\eta$  and the solution of the unperturbed ( $\epsilon = 0$ ) equation can be written as:

$$z(t) = z_1 e^{i\omega_1 t} + z_2 e^{-i\omega_2 t}, \quad z_1, z_2 \in \mathbb{C}, \quad (38)$$

with  $\omega_1 = \sqrt{\alpha^2 + 1} + \alpha, \omega_2 = \sqrt{\alpha^2 + 1} - \alpha$ . Applying variation of constants leads to equations for  $z_1$  and  $z_2$ :

$$\begin{aligned} \dot{z}_1 &= \frac{i\epsilon}{\omega_1 + \omega_2} \left( 2\kappa \left( i\omega_1 z_1 - i\omega_2 z_2 e^{-i(\omega_1 + \omega_2)t} \right) \right. \\ &\quad \left. + 4\eta^2 \cos 2\eta t \left( z_1 + z_2 e^{-i(\omega_1 + \omega_2)t} \right) \right), \\ \dot{z}_2 &= \frac{-i\epsilon}{\omega_1 + \omega_2} \left( 2\kappa \left( i\omega_1 z_1 e^{i(\omega_1 + \omega_2)t} - i\omega_2 z_2 \right) \right. \\ &\quad \left. + 4\eta^2 \cos 2\eta t \left( z_1 e^{i(\omega_1 + \omega_2)t} + z_2 \right) \right). \end{aligned} \quad (39)$$

To calculate the instability interval around the value  $\eta_0 = \frac{1}{2}(\omega_1 + \omega_2) = \sqrt{\alpha^2 + 1}$ , we apply perturbation theory to find for the stability boundary:

$$\begin{aligned}\eta_b &= \sqrt{1 + \alpha^2} \left( 1 \pm \varepsilon \sqrt{1 + \alpha^2 - \frac{\kappa^2}{\eta_0^2}} + \dots \right) \\ &= \sqrt{1 + \alpha^2} \left( 1 \pm \sqrt{(1 + \alpha^2)\varepsilon^2 - \left(\frac{\mu}{\eta_0}\right)^2} + \dots \right)\end{aligned}\quad (40)$$

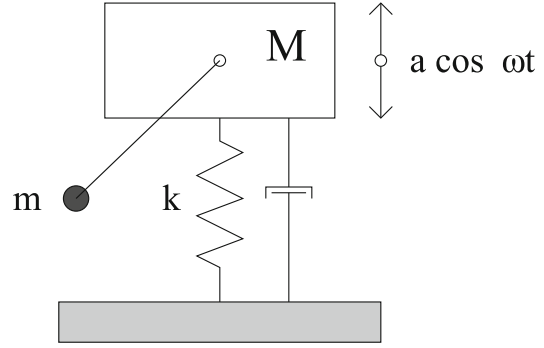
It follows that the domain of instability actually becomes *larger* when damping is introduced. The most unusual aspect of the above expression for the instability interval, however, is that there is a discontinuity at  $\kappa = 0$ . If  $\kappa \rightarrow 0$ , then the boundaries of the instability domain tend to the limits  $\eta_b \rightarrow \sqrt{1 + \alpha^2} (1 \pm \varepsilon \sqrt{1 + \alpha^2})$  which differs from the result we found when  $\kappa = 0$ :  $\eta_b = \sqrt{1 + \alpha^2} (1 \pm \varepsilon)$ .

In mechanical terms, the broadening of the instability-domain is caused by the coupling between the two degrees of freedom of the rotor in lateral directions which arises in the presence of damping. Such phenomena are typical for gyroscopic systems and have been noted earlier in the literature; see (Banichuk et al. 1989; Bolotin 1963) and (Seyranian and Mailybaev 2003). The explanation of the discontinuity and its genericity in (Hoveijn and Ruijgrok 1995), see subsection “Parametrically Forced Oscillators in Sum Resonance”, is new. For hysteresis and phase-locking phenomena in this problem, the reader is referred to (Ruijgrok et al. 1993).

### Autoparametric Excitation

In (Tondl et al. 2000), autoparametric systems are characterized as vibrating systems which consist of at least two consisting subsystems that are coupled. One is a Primary System that can be in normal mode vibration. In the instability (parameter) intervals of the normal mode solution in the full, coupled system, we have autoparametric resonance. The vibrations of the Primary System act as parametric excitation of the Secondary System which will no longer remain at rest. An example is presented in Fig. 5.

In actual engineering problems, we wish sometimes to diminish the vibration amplitudes of the Primary System; sometimes this is called ‘quenching of vibrations’. In other cases we have a coupled Secondary System which we would like



**Perturbation Analysis of Parametric Resonance, Fig. 5** Two coupled oscillators with vertical oscillations as Primary System and parametric excitation of the coupled pendulum (Secondary System)

to keep at rest. As an example we consider the following autoparametric system studied in (Fatimah and Ruijgrok 2002):

$$\begin{aligned}x'' + x + \varepsilon \left( k_1 x' + \sigma_1 x + a \cos 2\tau x + \frac{4}{3} x^3 + c_1 y^2 x \right) &= 0 \\ y'' + y + \varepsilon \left( k_2 y' + \sigma_2 y + c_2 x^2 y + \frac{4}{3} y^3 \right) &= 0\end{aligned}\quad (41)$$

where  $\sigma_1$  and  $\sigma_2$  are the detunings from the 1 : 1-resonance of the oscillators. In this system,  $y(t) = \dot{y}(t) = 0$  corresponds with a normal mode of the  $x$ -oscillator.

The system (41) is invariant under  $(x, y) \rightarrow (x, -y)$ ,  $(x, y) \rightarrow (-x, y)$ , and  $(x, y) \rightarrow (-x, -y)$ . Using the method of averaging as a normalization procedure we investigate the stability of solutions of system (41). To give an explicit example we follow (Fatimah and Ruijgrok 2002) in more detail. Introduce the usual variation of constants transformation:

$$\begin{aligned}x &= u_1 \cos \tau + v_1 \sin \tau ; \\ x' &= -u_1 \sin \tau + v_1 \cos \tau\end{aligned}\quad (42)$$

$$\begin{aligned}y &= u_2 \cos \tau + v_2 \sin \tau ; \\ y' &= -u_2 \sin \tau + v_2 \cos \tau\end{aligned}\quad (43)$$

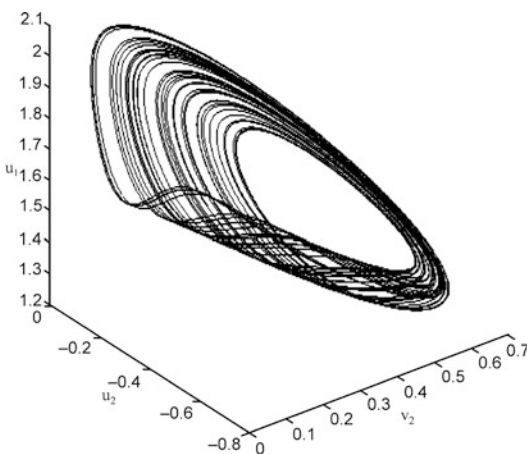
After rescaling  $\tau = \frac{\varepsilon}{2} \tilde{\tau}$  the averaged system of (41) becomes:



$$\begin{aligned}
 v_1' &= -k_1 v_1 - \left( \sigma_1 + \frac{1}{2}a \right) u_1 - u_1 (u_1^2 + v_1^2) \\
 &\quad - \frac{3}{4} c_1 u_2^2 u_1 - \frac{1}{4} c_1 v_2^2 u_1 - \frac{1}{2} c_1 u_2 v_2 v_1 \\
 u_2' &= -k_2 u_2 + \sigma_2 v_2 + v_2 (u_2^2 + v_2^2) + \frac{1}{4} c_2 u_1^2 v_2 \\
 &\quad + \frac{3}{4} c_2 v_1^2 v_2 + \frac{1}{2} c_2 u_1 v_1 u_2 \\
 v_2' &= -k_2 v_2 - \sigma_2 u_2 - u_2 (u_2^2 + v_2^2) - \frac{3}{4} c_2 u_1^2 u_2 \\
 &\quad - \frac{1}{4} c_2 v_1^2 u_2 - \frac{1}{2} c_2 u_1 v_1 v_2
 \end{aligned}
 \tag{44}$$

This system is analyzed for critical points, periodic and quasi-periodic solutions, producing existence and stability diagrams in parameter space. The system also contains a sequence of period-doubling bifurcations leading to chaotic solutions, see Fig. 6.

To prove the presence of chaos involves an application of higher dimensional Melnikov theory developed in (Wiggins 1988). A rather technical analysis in (Fatimah and Ruijgrok 2002) shows the existence of a Šilnikov orbit in the averaged equation, which implies chaotic dynamics, also for the original system.



**Perturbation Analysis of Parametric Resonance, Fig. 6** The strange attractor of the averaged system (44). The phase-portraits in the  $(u_2, v_2, u_1)$ -space for  $c_2 < 0$  at the value  $\sigma_2 = 5.3$ . The Kaplan–Yorke dimension for  $\sigma_2 = 5.3$  is 2.29

## Future Directions

Ongoing research in dynamical systems includes nonlinear systems with parametric resonance, but there are a number of special features as these systems are non-autonomous. This complicates the dynamics from the outset. For instance a two degrees of freedom system with parametric resonance involves at least three frequencies, producing many possible resonances. The analysis of such higher dimensional systems with many more combination resonances, has begun recently, producing interesting limit sets and invariant manifolds. Also the analysis of PDEs with periodic coefficients will play a part in the near future. These lines of research are of great interest.

In the conservative case, the association with Hamiltonian systems, KAM theory etc. gives a natural approach. This has already produced important results. In real-life modeling, there will always be dissipation and it is important to include this effect. Preliminary results suggest that the impact of damping on for instance quasi-periodic systems, is quite dramatic. This certainly merits more research.

Finally, applications are needed to solve actual problems *and* to inspire new, theoretical research.

**Acknowledgments** A number of improvements and clarifications were suggested by the editor, Giuseppe Gaeta. Additional references were obtained from Henk Broer and Fadi Dohnal.

## Bibliography

### Primary Literature

- Arnold VI (1983) Geometrical methods in the theory of ordinary differential equations. Springer, New York
- Banichuk NV, Bratus AS, Myshkis AD (1989) Stabilizing and destabilizing effects in nonconservative systems. PMM USSR 53(2):158–164
- Bogoliubov NN, Mitropolskii YA (1961) Asymptotic methods in the theory of nonlinear oscillations. Gordon and Breach, New York
- Bolotin VV (1963) Non-conservative problems of the theory of elastic stability. Pergamon Press, Oxford
- Broer HW, Levi M (1995) Geometrical aspects of stability theory for Hill's equation. Arch Ration Mech Anal 131: 225–240
- Broer HW, Simó C (1998) Hill's equation with quasi-periodic forcing: resonance tongues, instability pockets and global phenomena. Bol Soc Brasil Mat 29:253–293



- Broer HW, Simó C (2000) Resonance tongues in Hill's equations: a geometric approach. *J Differ Equ* 166: 290–327
- Broer HW, Vegter G (1992) Bifurcational aspects of parametric resonance, vol 1. In: Jones CKRT, Kirchgraber U, Walther HO (eds) *Expositions in dynamical systems*. Springer, Berlin, pp 1–51
- Broer HW, Hoveijn I, Van Noort M (1998) A reversible bifurcation analysis of the inverted pendulum. *Physica D* 112:50–63
- Broer HW, Hoveijn I, Van Noort M, Vegter G (1999) The inverted pendulum: a singularity theory approach. *J Diff Eqs* 157:120–149
- Broer HW, Puig J, Simó C (2003) Resonance tongues and instability pockets in the quasi-periodic Hill-Schrödinger equation. *Commun Math Phys* 241:467–503
- Broer HW, Hoveijn I, Van Noort M, Simó C, Vegter G (2005) The parametrically forced pendulum: a case study in  $1\frac{1}{2}$  degree of freedom. *J Dyn Diff Equat* 16: 897–947
- Cicogna G, Gaeta G (1999) Symmetry and perturbation theory in nonlinear dynamics. *Lecture Notes Physics*, vol 57. Springer, Berlin
- Fatimah S, Ruijgrok M (2002) Bifurcation in an auto-parametric system in 1:1 internal resonance with parametric excitation. *Int J Non-Linear Mech* 37:297–308
- Golubitsky M, Schaeffer D (1985) Singularities and groups in bifurcation theory. Springer, New York
- Hale J (1963) *Oscillation in nonlinear systems*. McGraw-Hill, New York, 1963; Dover, New York, 1992
- Hoveijn I, Ruijgrok M (1995) The stability of parametrically forced coupled oscillators in sum resonance. *ZAMP* 46:383–392
- Iooss G, Adelmeyer M (1992) *Topics in bifurcation theory*. World Scientific, Singapore
- Krupa M (1997) Robust heteroclinic cycles. *J Nonlinear Sci* 7:129–176
- Kuznetsov YA (2004) *Elements of applied bifurcation theory*, 3rd edn. Springer, New York
- Len JL, Rand RH (1988) Lie transforms applied to a nonlinear parametric excitation problem. *Int J Non-linear Mech* 23:297–313
- Levy DM, Keller JB (1963) Instability intervals of Hill's equation. *Commun Pure Appl Math* 16:469–476
- Magnus W, Winkler S (1966) *Hill's equation*. Interscience-John Wiley, New York
- McLaughlin JB (1981) Period-doubling bifurcations and chaotic motion for a parametrically forced pendulum. *J Stat Phys* 24:375–388
- Ng L, Rand RH (2002) Bifurcations in a Mathieu equation with cubic nonlinearities. *Chaos, Solitons Fractals* 14: 173–181
- Ng L, Rand RH (2003) Nonlinear effects on coexistence phenomenon in parametric excitation. *Nonlinear Dyn* 31:73–89
- Pikovsky AS, Feudel U (1995) Characterizing strange nonchaotic attractors. *Chaos* 5:253–260
- Ramani DV, Keith WL, Rand RH (2004) Perturbation solution for secondary bifurcation in the quadratically-damped Mathieu equation. *Int J Non-linear Mech* 39:491–502
- Recktenwald G, Rand RH (2005) Coexistence phenomenon in autoparametric excitation of two degree of freedom systems. *Int J Non-linear Mech* 40:1160–1170
- Roseau M (1966) *Vibrations nonlinéaires et théorie de la stabilité*. Springer, Berlin
- Ruijgrok M (1995) *Studies in parametric and autoparametric resonance*. Thesis, Utrecht University, Utrecht
- Ruijgrok M, Verhulst F (1996) Parametric and autoparametric resonance. *Prog Nonlinear Differ Equ Their Appl* 19:279–298
- Ruijgrok M, Tondl A, Verhulst F (1993) Resonance in a rigid rotor with elastic support. *ZAMM* 73:255–263
- Sanders JA, Verhulst F, Murdock J (2007) *Averaging methods in nonlinear dynamical systems*, rev edn. *Appl Math Sci*, vol 59. Springer, New York
- Seyranian AP (2001) Resonance domains for the Hill equation with allowance for damping. *Physics* 46: 41–44
- Seyranian AP, Mailybaev AA (2003) *Multiparameter stability theory with mechanical applications*. Series A, vol 13. World Scientific, Singapore
- Seyranian AA, Seyranian AP (2006) The stability of an inverted pendulum with a vibrating suspension point. *J Appl Math Mech* 70:754–761
- Stoker JJ (1950) *Nonlinear vibrations in mechanical and electrical systems*. Interscience, New York, 1950; Wiley, New York, 1992
- Strutt MJO (1932) *Lamé-sche, Mathieu-sche und verwandte Funktionen*. Springer, Berlin
- Szemplinska-Stupnicka W (1990) The behaviour of nonlinear vibrating systems, vol 2. Kluwer, Dordrecht
- Tondl A (1991a) *Quenching of self-excited vibrations*. Elsevier, Amsterdam
- Tondl A, Ruijgrok M, Verhulst F, Nabergoj R (2000) *Autoparametric resonance in mechanical systems*. Cambridge University Press, New York
- Van der Pol B, Strutt MJO (1928) On the stability of the solutions of Mathieu's equation. *Phil Mag Lond Edinb Dublin* 7(5):18–38
- Verhulst F (1996) *Nonlinear differential equations and dynamical systems*. Springer, New York
- Verhulst F (2005a) Invariant manifolds in dissipative dynamical systems. *Acta Appl Math* 87:229–244
- Verhulst F (2005b) *Methods and applications of singular perturbations*. Springer, New York
- Wiggins S (1988) *Global bifurcation and chaos*. *Appl Math Sci*, vol 73. Springer, New York
- Yakubovich VA, Starzhinskii VM (1975) *Linear differential equations with periodic coefficients*, vol 1 and 2. Wiley, New York
- Zounes RS, Rand RH (1998) Transition curves for the quasi-periodic Mathieu equation. *SIAM J Appl Math* 58:1094–1115
- Zounes RS, Rand RH (2002) Global behavior of a nonlinear quasi-periodic Mathieu equation. *Nonlinear Dyn* 27:87–105

## Books and Reviews

- Arnold VI (1977) Loss of stability of self-oscillation close to resonance and versal deformation of equivariant vector fields. *Funct Anal Appl* 11:85–92
- Arscott FM (1964) Periodic differential equations. Macmillan, New York
- Cartmell M (1990) Introduction to linear, parametric and nonlinear vibrations. Chapman and Hall, London
- Dohnal F (2005) Damping of mechanical vibrations by parametric excitation. PhD thesis, Vienna University of Technology
- Dohnal F, Verhulst F (2008) Averaging in vibration suppression by parametric stiffness excitation. *Nonlinear Dyn* (accepted for publication)
- Ecker H (2005) Suppression of self-excited vibrations in mechanical systems by parametric stiffness excitation. *Fortschrittsberichte Simulation Bd 11*. Argesim/Asim Verlag, Vienna
- Fatimah S (2002) Bifurcations in dynamical systems with parametric excitation. Thesis, University of Utrecht
- Hale J (1969) Ordinary differential equations. Wiley, New York
- Kirillov ON (2007) Gyroscopic stabilization in the presence of nonconservative forces. *Dokl Math* 76:780–785; Orig Russian: (2007) *Dokl Ak Nauk* 416:451–456
- Meixner J, Schäfer FW (1954) Mathieu'sche Funktionen und Sphäroidfunktionen. Springer, Berlin
- Moon FC (1987) Chaotic vibrations: an introduction for applied scientists and engineers. Wiley, New York
- Nayfeh AH, Mook DT (1979) Nonlinear Oscillations. Wiley Interscience, New York
- Schmidt G (1975) Parametererregte Schwingungen. VEB Deutscher Verlag der Wissenschaften, Berlin
- Schmidt G, Tondl A (1986) Non-linear vibrations. Akademie-Verlag, Berlin
- Tondl A (1978) On the interaction between self-excited and parametric vibrations. In: Monographs and memoranda, vol 25. National Res Inst Bechovice, Prague
- Tondl A (1991b) On the stability of a rotor system. *Acta Technica CSAV* 36:331–338
- Tondl A (2003) Combination resonances and anti-resonances in systems parametrically excited by harmonic variation of linear damping coefficients. *Acta Technica CSAV* 48:239–248
- Van der Burgh AHP, Hartono AAK (2004) Rain-wind induced vibrations of a simple oscillator. *Int J Non-Linear Mech* 39:93–100
- Van der Burgh AHP, Hartono AAK (2006) A new model for the study of rain-wind-induced vibrations of a simple oscillator. *Int J Non-Linear Mech* 41:345–358
- Weinstein A, Keller JB (1985) Hill's equation with a large potential. *SIAM J Appl Math* 45:200–214
- Weinstein A, Keller JB (1987) Asymptotic behaviour of stability regions for Hill's equation. *SIAM J Appl Math* 47:941–958
- Wiggins S (1990) Introduction to applied nonlinear dynamical systems and chaos. Springer, New York

# Symmetry and Perturbation Theory in Non-linear Dynamics

Giuseppe Gaeta  
Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Symmetry of Dynamical Systems  
Perturbation Theory: Normal Forms  
Perturbative Determination of Symmetries  
Symmetry Characterization of Normal Forms  
Symmetries and Transformation to Normal Form  
Generalizations  
Symmetry for Systems in Normal Form  
Linearization of a Dynamical System  
Further Normalization and Symmetry  
Symmetry Reduction of Symmetric Normal Forms  
Conclusions  
Future Developments  
Additional Notes  
Bibliography

## Glossary

**Perturbation theory** A theory aiming at studying solutions of a differential equation (or system thereof), possibly depending on external parameters, near a known solution and/or for values of external parameters near to those for which solutions are known.

**Dynamical system** A system of first order differential equations  $dx^i/dt = f^i(x, t)$ , where  $x \in M$ ,  $t \in \mathbf{R}$ . The space  $M$  is the phase space for the dynamical system, and  $\tilde{M} = M \times \mathbf{R}$  is the extended phase space. When  $f$  is

smooth we say the dynamical system is smooth, and for  $f$  independent of  $t$ , we speak of an autonomous dynamical system.

**Symmetry** An invertible transformation of  $\tilde{M}$  mapping solutions into solutions. If the dynamical system is smooth, smoothness will also be required on symmetry transformations; if it is autonomous, it will be natural to consider transformations of  $M$  rather than of  $\tilde{M}$ .

**Symmetry reduction** A method to reduce the equations under study to simpler ones (e.g. with less dependent variables, or of lower degree) by exploiting their symmetry properties.

**Normal form** A convenient form to which the system of differential equations under study can be brought by means of a sequence of change of coordinates. The latter are in general well defined only in a subset of  $M$ , possibly near a known solution for the differential equations.

**Further normalization** A procedure to further simplify the normal form for a dynamical system, in general making use of certain degeneracies in the equations to be solved in the course of the normalization procedure.

## Definition of the Subject

Given a differential equation or system of differential equations  $\Delta$  with independent variables  $\xi^a \in \Xi \subseteq \mathbf{R}^q$  and dependent variables  $x^i \in M \subseteq \mathbf{R}^p$ , a symmetry of  $\Delta$  is an invertible transformation of the extended phase space  $\tilde{M} = \Xi \times M$  into itself which maps solutions of  $\Delta$  into (generally, different) solutions of  $\Delta$ .

The presence of symmetries is a non-generic feature; correspondingly, equations with symmetry have some special features. These can be used to obtain information about the equation and its solutions, and sometimes allow one to obtain explicit solutions.

The same applies when we consider a perturbative approach to the equations: taking

into account the presence of symmetries guarantees the perturbative expansion has certain specific features (e.g. some terms are not allowed) and hence allows one to deal with simplified expansions and equations; thus this approach can be of great help in providing explicit solutions.

As mentioned above, symmetry is a non-generic feature: if we take a “generic” equation or system, it will not have any symmetry property. What makes the symmetry approach useful and widely applicable is a remarkable fact: many of the equations encountered in applications, and especially in physical and related ones (mechanical, electronic, etc.) *are* symmetric; this in turn descends from the fact that the fundamental equations of physics have a high degree of symmetry.

Thus, symmetry-based methods are at the same time “non-generic” in a mathematical sense, and “general” in a physical, or more generally real-world, sense.

## Introduction

Symmetry has been a major ingredient in the development of quantum perturbation theory, and is a fundamental ingredient of the theory of integrable (Hamiltonian and non-Hamiltonian) systems; yet, the use of symmetry in the context of general perturbation theory is rather recent.

From the point of view of nonlinear dynamics, the use of symmetry has become widespread only through equivariant bifurcation theory; even in this case, attention has been mostly confined to linear symmetries.

Also, in recent years the theory and practice of symmetry methods for differential equations became increasingly popular and has been applied to a variety of problems (to a large extent, following the appearance of the book by Olver (1986)). This theory is deeply geometrical and deals with symmetries of general nature (provided that they are described by smooth vector fields), i.e. in this context there is no reason to limit attention to linear symmetries.

In this article we look at the basic tools of perturbation theory, i.e. *normal forms* (first introduced by Poincaré more than a century ago for

general dynamical systems; the Hamiltonian case being studied in its special features by Birkhoff several decades ago) and study their interaction with symmetries, with no limitation to linear ones. See the entries ► [“Normal Forms in Perturbation Theory”](#), ► [“Hamiltonian Perturbation Theory \(and Transition to Chaos\)”](#) for an introduction to Normal Forms.

We focus on the most basic setting, i.e. systems having a fixed point (at the origin) and perturbative expansions around this; thus our theory is entirely local. We also limit to the discussion of general vector fields, i.e. we will not discuss the formulation one would obtain for the special case of Hamiltonian vector fields ► [“Hamiltonian Perturbation Theory \(and Transition to Chaos\)”](#), (Hanssmann 2007) (in which case one can deal with the Hamiltonian function rather than with the vector field it generates), referring the reader to (Cicogna and Gaeta 1999) for this as well as for other extensions and for several proofs.

We start by recalling basic notions about the symmetry of differential equations, and in particular of dynamical systems; we will then discuss normal forms in the presence of symmetries, and the problem of taking into normal form the dynamical vector field and the symmetry vector field(s) at the same time.

The presence of symmetry causes several peculiar phenomena in the dynamics, and hence also in perturbative expansions. This has been explained in very effective terms by Ian Stewart (1988):

Symmetries abound in nature, in technology, and?—especially—in the simplified mathematical models we study so assiduously. Symmetries complicate things and simplify them. They complicate them by introducing exceptional types of behavior, increasing the number of variables involved, and making vanish things that usually do not vanish. They simplify them by introducing exceptional types of behavior, increasing the number of variables involved, and making vanish things that usually do not vanish. They violate all the hypotheses of our favorite theorems, yet lead to natural generalizations of those theorems. It is now standard to study the “generic” behavior of dynamical systems. Symmetry is not generic. The answer is to work within the world of symmetric systems and to examine a suitably restricted idea of genericity.

Here we deal with dynamical systems, and more specially autonomous ones, i.e. systems of

equations of the form  $dx^i/dt = f^i(x)$ . Now we have a single independent variable, the time  $t \in \mathbf{R}$ , and in view of its distinguished role we will mainly focus attention on transformations leaving it unchanged.

It is appropriate to point out here connections to several topics which we will *not* illustrate in this article.

First of all, we stress that we will work at the formal level, i.e. without considering the problem of convergence of the power series entering in the theory. This convergence is studied in the articles ► “[Perturbation Theory](#)”, ► “[Convergence of Perturbative Expansions](#)”, to which the interested reader is referred in the first instance.

As hinted above, perturbation theory for symmetric systems has many points of contact with the topic of Equivariant Bifurcation Theory, which we will not touch upon here. The interested reader is referred to (Glendinning 1994; Guckenheimer and Holmes 1983; Iooss and Adelmeyer 1992; Ruelle 1989; Verhulst 1989) for Bifurcation Theory in general, and then for the equivariant setting to the books (Chossat and Lauterbach 1999; Golubitsky et al. 1988; Iooss and Adelmeyer 1992). More compact introductions are provided by the review papers (Crawford 1991; Gaeta 1990).

Many facets of the interplay of symmetry and perturbation theory are also discussed in the SPT conference proceedings volumes (Abenda et al. 2003; Bambusi and Gaeta 1997; Bambusi et al. 2001; Degasperis and Gaeta 1999; Gaeta et al. 2005, 2007).

Our discussion is based on the treatment in (Cicogna and Gaeta 1999), with integrations and updates where appropriate.

Some considerations and remarks are given in additional notes collected in the last section; these are called for by marks<sup>(xx)</sup> with  $xx$  consecutive numbers.

## Symmetry of Dynamical Systems

Symmetry of differential equations – and its use to solve or reduce the differential equations themselves – is a classical and venerable subject, being

the very motivation to Sophus Lie when he created what is nowadays known as the theory of Lie groups (Kirillov 1976). The subject is now dealt with in a number of textbooks (see e.g. (Aleekseevskij et al. 1991; Baumann 2000; Bluman and Anco 2002; Bluman and Kumei 1989; Cantwell 2002; Gaeta 1994; Hydon 2000; Krasil’shchik and Vinogradov 1999; Olver 1986, 1995; Stephani 1989)) and review papers (see e.g. (Ibragimov 1992; Vorob’ev 1986, 1991; Winternitz 1987, 1993)); we will thus refer the reader to these for the general theory, and briefly recall here the special formulation one obtains when dealing with symmetries of smooth dynamical systems in  $\mathbf{R}^n$ .

Consider a (possibly non-autonomous) system

$$\dot{x}^i = f^i(x; t) \quad i = 1, \dots, n; \quad (1)$$

we assume  $x \in M = \mathbf{R}^n$ ;  $M$  is also called the phase space, and  $\tilde{M} = M \times \mathbf{R}$  (the second factor representing of course time  $t$ ) is the extended phase space.<sup>(1)</sup>

We consider now vector fields in  $\tilde{M}$ ; these can be written in coordinates as

$$S = \tau(x, t) \frac{\partial}{\partial t} + \sum_{i=1}^n \varphi^i(x, t) \frac{\partial}{\partial x^i}. \quad (2)$$

Note that (1) is identified with the vector field

$$X_f := \sum_{i=1}^n f^i(x, t) \frac{\partial}{\partial x^i}. \quad (3)$$

A (vector) function  $x : \mathbf{R} \rightarrow M$  is naturally identified with the subset  $\sigma_x$  of  $\tilde{M}$  (corresponding to its graph) defined by

$$\sigma_x = \{(y, t) \in M \times \mathbf{R} : y^i = x^i(t)\} \subset \tilde{M}. \quad (4)$$

The vector field  $S$  acts infinitesimally in  $\tilde{M}$  by mapping points  $(y, t)$  to points  $(\hat{y}, \hat{t})$  given by

$$\hat{t} = t + \varepsilon \tau(y, t), \quad \hat{y}^i = y^i + \varepsilon \varphi^i(y, t); \quad (5)$$

as  $\varepsilon$  is small these relations can be inverted, yielding at first order in  $\varepsilon$

$$t = \hat{t} - \varepsilon \tau(\hat{y}, \hat{t}), \quad y^i = \hat{y}^i - \varepsilon \varphi^i(\hat{y}, \hat{t}). \quad (6)$$

Using these relations, it is easy to check that the subset  $\sigma = \sigma_x$  is mapped by  $S$  to a (generally) different subset  $\hat{\sigma}$ , corresponding to  $y = \hat{x}(t)$ , with

$$\hat{x}^j(t) = x^j(t) + \varepsilon [\varphi^j(x(t), t) - \dot{x}^j(t) \tau(x(t), t)]. \quad (7)$$

We say that  $S$  is a **symmetry** for the dynamical system (1) if it maps solutions into (generally, different) solutions. The condition for this to happen turns out to be (Cicogna and Gaeta 1999)

$$\frac{\partial \varphi^i}{\partial t} - \frac{\partial \tau}{\partial t} f^i = \varphi^j \frac{\partial f^i}{\partial x^j} - f^j \frac{\partial \varphi^i}{\partial x^j}. \quad (8)$$

This can be more compactly expressed by introducing the Lie–Poisson bracket

$$\{f, g\} := (f \cdot \nabla) g - (g \cdot \nabla) f \quad (9)$$

between vector functions on  $\tilde{M}$ . Then (8) reads

$$(\partial \varphi / \partial t) - (\partial \tau / \partial t) f = \{f, \varphi\}. \quad (10)$$

In the following we will consider autonomous dynamical systems; in this case it is rather natural to consider only transformations which leave  $t$  invariant, i.e. with  $\tau = 0$ . In this case (10) reduces to

$$(\partial \varphi / \partial t) + \{f, \varphi\} = 0. \quad (11)$$

A further reduction is obtained if we only consider transformations for which the action on  $M$  is also independent of time, so that  $\partial \varphi / \partial t = 0$  and the symmetry condition is

$$\{f, \varphi\} = 0; \quad (12)$$

in this case one speaks of *Lie-point time-independent (LPTI) symmetries*.

The Eqs. (8) (or its reductions) will be referred to as the determining equations for the symmetries of the dynamical system (1).<sup>(2)</sup>

It should be stressed that (8) are linear in  $\varphi$  and  $\tau$ ; it is thus obvious that the solutions will span a linear space. It is also easy to check (the proof of this fact follows from the bilinearity of (9) and the Jacobi identity) that if  $S_1$  and  $S_2$  are solutions

to (8), so is their Lie–Poisson bracket  $\{S_1, S_2\}$ . The set  $\mathcal{G}_{X_f}$  of vector fields  $X_\varphi$  with  $\varphi$  solutions to (8) is thus a Lie algebra; it is the *symmetry algebra* for the dynamical system (1).

The symmetry algebra of a dynamical system is infinite dimensional, but has moreover an additional structure. That is, it is a module over the algebra  $\mathcal{I}_{X_f}$  of first integrals for  $f$  (that is, scalar functions  $\alpha : M \rightarrow \mathbf{R}$  such that  $X_f(\alpha) \equiv (f \cdot \nabla)\alpha = 0$ ). Albeit  $\mathcal{G}_{X_f}$  is infinite dimensional as a Lie algebra, it is not so as a Lie module. We have, indeed (Walcher 1991):

**Theorem 1 (Walcher)** The set  $\mathcal{G}_{X_f}$  is a finitely generated module over  $\mathcal{I}_{X_f}$ .

## Perturbation Theory: Normal Forms

In this section we recall some basic facts about perturbation theory for general dynamical systems, referring to ► “Normal Forms in Perturbation Theory”, ► “Hamiltonian Perturbation Theory (and Transition to Chaos)”, ► “Perturbation Theory” for details. For the sake of simplicity, we discuss perturbations around an equilibrium point; see e.g. (Arnold 1976, 1980; Arnold and Il’yashenko 1988; Hanssmann 2007; Ruelle 1989; Sanders and Verhulst 1985; Sanders et al. 2007) for more general settings.

As is well known – and discussed, e.g. in ► “Normal Forms in Perturbation Theory”, ► “Hamiltonian Perturbation Theory (and Transition to Chaos)”, – a central objective of perturbation theory is to set (1) in normal form, i.e. to eliminate as many nonlinear terms as possible, so that the difference with respect to the linearized equation is as small as possible (see again ► “Normal Forms in Perturbation Theory”, ► “Hamiltonian Perturbation Theory (and Transition to Chaos)”, or ► “Perturbation Theory”, for a precise meaning to this statement; a lengthier discussion is given e.g. in (Arnold 1974, 1980; Gallavotti 1983; Glendinning 1994; Verhulst 1989)).<sup>(3)</sup>

We briefly recall how this goes, also in order to fix notation. We consider a  $C^\infty$  dynamical system  $\dot{x} = f(x)$  in  $\mathbf{R}^n$ , admitting  $x = 0$  as an equilibrium point – that is with  $f(0) = 0$ . By Taylor-expanding  $f(x)$  around  $x = 0$  we will write this in the form

$$\dot{x} = f(x) = \sum_{k=0}^{\infty} F_k(x) \quad (13)$$

where  $F_k(x)$  is homogeneous of degree  $(k+1)$  in  $x$  (this seemingly odd notation will come out handy in the following). We denote the linear space of vector function  $f: \mathbf{R}^n \rightarrow \mathbf{R}^n$  homogeneous of degree  $(k+1)$  by  $\mathcal{V}_k$ .

### Poincaré–Dulac Normal Forms

Let us consider a change of coordinates of the form

$$\xi^i = x^i + h_k^i(x), \quad (14)$$

where  $h_k \in \mathcal{V}_k$  ( $k \geq 1$ ); we write  $F_j^i = (\partial h^i / \partial x^j)$ . A change of coordinates of the form (14) is called a Poincaré transformation (or P transformation for short), and the function  $h_k$  is also called the *generator* of the P transformation. In the following we will freely drop the subscript  $k$  when this cannot generate any confusion.

The transformation (14) is, for small  $x$ , a near-identity transformation; thus it is surely invertible in a small enough neighborhood of the origin. We apply  $A := (I + \Gamma)^{-1}$  on (13), and get the P transformed dynamical system in the form

$$\dot{x} = \tilde{f}(x) \equiv \Lambda \sum_{m=0}^{\infty} F_m(x + h_k(x)) = \sum_{m=0}^{\infty} \tilde{F}_m(x). \quad (15)$$

In order to identify the  $\tilde{F}_m$  we should consider power series expansions for  $A$  and for  $F_m(x + h(x))$ . With standard computations (we refer again to ► “Normal Forms in Perturbation Theory”, or to (Arnold 1980; Cicogna and Gaeta 1999), for details), we obtain that the  $\tilde{F}_m$  are given (with  $[q]$  the integer part of  $q$ ) by

$$\tilde{F}_m = F_m + \sum_{p=1}^{[m/k]} \left[ \sum_{s=0}^p (-1)^s \Gamma^s \Phi_{h_k}^{p-s} \right] F_{m-kp}. \quad (16)$$

The  $\Phi_h^r$  appearing in (16) are defined as follows. With a multi-index notation, write  $J = (j_1, \dots, j_n)$ ,  $|J| = \sum_i j_i$ ; set then  $\partial_J := \partial_1^{j_1} \dots \partial_n^{j_n}$ ,

and similarly  $h_k^J := (h_k^1)^{j_1} \dots (h_k^n)^{j_n}$ . The operators  $\Phi_h^r$  (representing all the partial derivatives of order  $|J|$ ) are defined as  $\Phi_h^r = (1/r!) \sum (h^J \cdot \partial_J)$ .

Some special cases following from this general formula should be noted. As well known, the terms of degree smaller than  $k$  are not changed at all, i.e.  $\tilde{F}_m = F_m$  for  $m < k$ , and the term of degree  $k$  is changed (writing  $h \equiv h_k$ ) according to

$$\tilde{F}_k = F_k + [\Phi_h - \Gamma]F_0. \quad (17)$$

(Similarly, for  $0 < v < k$ , the term of degree  $k+v$  is changed into  $\tilde{F}_{k+v} = F_{k+v} + [\Phi_h - \Gamma]F_v$ .) Define now, recalling (9), the operators

$$L_k = \{F_k, \cdot\}; \quad (18)$$

note  $L_k: \mathcal{V}_m \rightarrow \mathcal{V}_{m+k}$ .

The operator  $\mathcal{A} = L_0$ , associated with the linear part  $A = (Df)(0)$  of  $f$  (that is,  $F_0(x) = Ax^{(4)}$ ) is called the homological operator; it leaves the spaces  $\mathcal{V}_k$  invariant, and hence it admits the decomposition  $\mathcal{A} = \bigoplus_{m=0}^{\infty} \mathcal{A}^{(m)}$ , where  $\mathcal{A}^{(m)}$  is just the restriction of  $\mathcal{A}$  to  $\mathcal{V}_m$ .

In the following, we will need to consider the adjoint  $\mathcal{A}^+$  of the operator  $\mathcal{A}$ . For this we need to introduce a scalar product in the space  $\mathcal{V} = \bigcup \mathcal{V}_k$ . Actually, we can introduce a scalar product in each of the spaces  $\mathcal{V}_k$  into which  $\mathcal{V}$  decomposes.

A convenient scalar product was introduced in (Elphick et al. 1987) (following (Bargmann 1961)); we will only use this one (the reader should be warned that different definitions are also considered in the literature (Arnold 1974; Cicogna and Gaeta 1999)). We denote by  $\xi_{\mu, i}$  the vector function whose components are all zero but the  $i$ th one, given by  $x^{\mu} := x_1^{\mu_1} \dots x_n^{\mu_n}$ ; with this notation, we define<sup>(5)</sup>

$$(\xi_{\mu, i}, \xi_{\nu, j}) = \sum_{i=1}^n \langle \xi_{\mu, i}, \xi_{\nu, i} \rangle, \quad (19)$$

where  $\langle x^{\mu}, x^{\nu} \rangle \equiv \partial^{\mu} x^{\nu}$ , and  $\partial^{\mu} = \partial_1^{\mu_1} \dots \partial_n^{\mu_n}$ . (When in the following we consider adjoint operators, these will be understood in terms of this.)

With this scalar product, one has the following lemma (a proof is given, e.g. in (Iooss and Adelmeyer 1992)).



**Lemma 1** If  $\mathcal{A}$  is the homological operator associated with the matrix  $A$ ,  $\mathcal{A} = \{Ax, \cdot\}$ , then its adjoint  $\mathcal{A}^+$  is the homological operator associated with the adjoint matrix  $A^+$ , i.e.,  $\mathcal{A}^+ = \{A^+x, \cdot\}$ .

We will also consider the projection onto the range of  $\mathcal{A}^{(k)}$ , denoted by  $\pi_k$ . The *general homological equations* (the one for  $k = 0$  corresponds to the standard homological equation) are then

$$\mathcal{A}^{(k)}(h_k) = \pi_k F_k. \quad (20)$$

These are equations for  $h_k \in \mathcal{V}_k$ , and always admit a solution (thanks to the presence of the projection operator  $\pi_k$ ). The Eq. (20) maps into a set of algebraic equations once we introduce a basis in  $\mathcal{V}_k$ .

The  $h_k \in \mathcal{V}_k$  solving to (20) will be of the form  $h_k = h_k^* + \ell_k$ , where  $h_k^* = \mathcal{A}^*(\pi_k F_k) \in \text{Ran}[(\mathcal{A}^{(k)})^*]$  (here  $\mathcal{A}^*$  is the pseudo-inverse to  $\mathcal{A}$ ; note  $\text{Ker}(\mathcal{A}^*) = \text{Ker}(\mathcal{A}^+)$ ) is unique, and  $\ell_k$  is any function in  $\text{Ker}[\mathcal{A}^{(k)}]$ .

**Remark 1** It should be stressed that, while adding a nonzero  $\ell_k$  to  $h_k^*$  does not change the resulting  $\tilde{F}_k$ , it could – and in general will – affect the terms of higher order.

One can then normalize  $X_f$  in the standard recursive way, based on solving homological equations; this is described, e.g. in ► “Normal Forms in Perturbation Theory”. In this way, we are reduced to considering only systems with

$$F_k \in [\text{Ran}(\mathcal{A}^{(k)})]^\perp = \text{Ker}[(\mathcal{A}^{(k)})^+]. \quad (21)$$

Such terms are also called *resonant*.<sup>(6)</sup>

The presence of resonant terms is related to the existence of resonance relations among eigenvalues  $\lambda_i$  of the matrix  $A$  describing the linear part of the system; these are relations of the form  $(m \cdot \lambda) \equiv \sum_i m_i \lambda_i = \lambda_s$  where the  $m_i$  are non-negative integers, with  $|m| = \sum m_i > 1$  (the restriction  $|m| > 1$  is to avoid trivial cases); the integer  $|m|$  is also called the *order* of the resonance (► “Normal Forms in Perturbation Theory”, (Arnold 1980; Arnold and Il’yashenko 1988)).

If the system  $\dot{x} = f(x)$  has only resonant non-linear terms, we say that it is in Poincaré–Dulac normal form (► “Normal Forms in Perturbation Theory”, (Arnold 1980; Arnold and Il’yashenko 1988; Bryuno 1989; Chow et al. 1994)). If all the nonlinear terms of order up to  $q$  are resonant, we say that the system is in normal form up to order  $q$ .

**Theorem 2 (Poincaré–Dulac)** Any analytic vector field  $X_f$  with  $f(0) = 0$  can be formally taken to normal form by a sequence of Poincaré transformations.

**Remark 2** If we do not have an exact resonance, but  $(m \cdot \lambda) - \lambda_s \simeq 0$ , we have a small denominator, and correspondingly a very large coefficient in  $h_k$ , where  $k = |m|$ . Such small denominators are responsible for divergencies in the normalizing series (Arnold 1980; Arnold and Il’yashenko 1988; Arnold et al. 1993; Bryuno 1971a, b, 1989; Siegel and Moser 1971).

### Lie Transforms

In discussing Poincaré normal forms, we have considered near-identity diffeomorphisms of  $M$ ; these can be expressed as time-one maps for the flow under some vector fields  $X_h$ . In a number of cases, it is more convenient to deal directly with such vector fields. We are going to briefly discuss this approach, and its relation with the one discussed above; for further detail, see e.g. (► “Normal Forms in Perturbation Theory”, (Benettin et al. 1984; Bogoliubov and Mitropolsky 1961; Broer 1979, 1981; Cicogna and Gaeta 1999; Fassò 1990; Gaeta 2001; Giorgilli and Locatelli 1997; Mitropolsky and Lopatin 1995)).

Let the vector field  $H \equiv X_h$  be given, in the  $x$  coordinates, by  $H = h^i(x)(\partial/\partial x^i)$ . We denote by  $\Psi(s; x)$  the local flow under  $H$  starting at  $x$ , so that  $(d/ds)\Psi(s; x) = H(\Psi(s; x))$ . We also use exponential notation:  $\Psi(s; x) = e^{sH}x$ .

We will denote  $\Psi(1; \xi)$  as  $x$ ; the direct and inverse changes of coordinates will be defined as

$$\begin{aligned} x &= \Psi(1; \xi) = [e^{sH}\xi]_{s=1}, \\ \xi &= \Psi(-1; x) = [e^{-sH}x]_{s=1}. \end{aligned} \quad (22)$$

Now, consider another vector field  $X$  on  $M$ , describing the dynamical system we are interested

in. If we study the dynamical system  $\dot{\xi}^i = f^i(\xi)$ , we consider the vector field given in the  $x$  coordinates by

$$X = X_f = f^i(x) (\partial / \partial x^i). \quad (23)$$

This also generates a (local) flow, i.e. for any  $\xi_0 \in M$  we have a one-parameter family  $\xi(t) \equiv \Gamma(t; \xi_0) \in M$  such that  $(d\xi(t)/dt) = X(\xi)$ . By means of (22), this also defines a one-parameter family  $x(t) \in M$ , which will satisfy  $(dx(t)/dt) = \tilde{X}(x)$  for some vector field  $\tilde{X}$  on  $M$ ; this will be the transformed vector field under (22), and is given by<sup>(7)</sup>

$$\tilde{X} = [e^{sH} X e^{-sH}]_{(s=1)} \quad (24)$$

We call this transformation the **Lie-Poincaré transformation** generated by  $h$ . Notice that this yields, up to order one in  $s$ , (and therefore if  $h \in \mathcal{V}_k$ , up to terms in  $\mathcal{V}_k$ ), just the same result as the Poincaré transformation with the same generator  $h$ .

The  $\tilde{X}$  can be given (for arbitrary  $s$ ), in terms of the Baker–Campbell–Hausdorff formula, as

$$\tilde{X} = \sum_{k=0}^{\infty} \frac{(-1)^k s^k}{k!} X^{(k)}, \quad (25)$$

where  $X^{(0)} = X$ , and  $X^{(k+1)}$  are defined recursively by  $X^{(k+1)} = [X^{(k)}, H]$ .

## Perturbative Determination of Symmetries

Let us now consider the problem of determining the symmetries of a given dynamical system. Writing the latter in the form

$$\dot{x} = f(x) = \sum_{k=0}^{\infty} F_k(x) \quad (F_k \in \mathcal{V}_k), \quad (26)$$

it is quite natural to also look for symmetries in terms of a (possibly only formal) power series; this will be our approach here.

Consider a generic smooth vector field (by smooth we mean, here and elsewhere,  $C^\infty$ ;

however, we will soon go on to actually consider vector fields represented by power series) in  $M \times \mathbf{R}$ ,

$$Y = \tau(x, t) \partial_t + \sum_{i=1}^n \varphi^i(x, t) \partial_i. \quad (27)$$

Here and below,  $\partial_i \equiv (\partial / \partial x^i)$ . As remarked in section “[Symmetry of Dynamical Systems](#)”, (27) is too general for our purposes; we are primarily interested in vector fields acting on  $M$  alone, and mainly (Cicogna and Gaeta 1994a, b, 1999; Gaeta 1994) in time-independent vector fields on  $M$  (note in this way the dynamical and the symmetry vector fields are on the same footing). Thus we will just consider (we will consistently use  $X$  for the vector field  $X_f$  defined by (26), and  $Y$  for the symmetry vector field  $X_s$ , in order to simplify the notation)

$$Y = \sum_{i=1}^n s^i(x) \partial_i. \quad (28)$$

Note that at this stage we are **not** assuming the dynamical system described by  $X$  has been taken into normal form; we will see later on the specific features of this case.

## Determining Equations

Now, let us consider the dynamical system identified by  $X$ , and look for the determining equation identifying its symmetries  $Y$  as in (28). Condition (8) yields

$$\begin{aligned} f^j(x) \frac{\partial s^i(x)}{\partial x^j} - s^j(x) \frac{\partial f^i(x)}{\partial x^j} \\ \equiv (f \cdot \nabla) s - (s \cdot \nabla) f = 0. \end{aligned} \quad (29)$$

As discussed above, using (9) this is also written as  $\{f, s\} = 0$ , which just means (as had to be expected)

$$[X, Y] = 0. \quad (30)$$

We now denote the set of  $Y$  satisfying (30) by  $\mathcal{G}_X$ . It is obvious that  $\mathcal{G}_X$ , equipped with the usual commutator of vector fields, is a Lie algebra. It is also easy to see that, as  $f(0) = 0$ ,  $Y \in \mathcal{G}_X$  implies  $s(0) = 0$ .

We can now expand  $Y$ , i.e.  $s(x)$ , in a perturbative series around  $x_0 = 0$  in the same way as we did for  $X$ . We write

$$s(x) = \sum_{k=0}^{\infty} S_k(x) \quad (S_k \in \mathcal{V}_k). \quad (31)$$

Plugging this into the determining Eqs. (30) we get, after rearranging the terms,

$$\sum_{k=0}^{\infty} \sum_{m=0}^k \{F_m, S_{k-m}\} = 0. \quad (32)$$

For this to hold, the different homogeneous terms of degree  $k$  must vanish separately. Thus, we have a hierarchy (in a sense to be explained in a moment) of equations

$$\sum_{m=0}^k \{F_m, S_{k-m}\} = 0 \quad k = 0, 1, 2, 3, \dots \quad (33)$$

It is convenient to isolate the terms containing linear factors, i.e. to rewrite (33) – for  $k \geq 1$  – in the form

$$\begin{aligned} \{F_0, S_k\} - \{S_0, F_k\} &= - \sum_{m=1}^{k-1} \{F_m, S_{k-m}\} \\ &\equiv \chi_k, \end{aligned} \quad (34)$$

where we have used the antisymmetry of  $\{., .\}$ , and  $\chi_0 = \chi_1 = 0$ .

### Recursive Solution of the Determining Equations

Let us now consider the problem of concretely solving the determining Eq. (32). As the perturbative series expansion suggests, we can proceed order by order, i.e. start with consideration of the equation for  $k = 0$ , then tackle  $k = 1$ , and so on.

Proceeding in this way we are always reduced to consider equations of the form

$$F_0^j \left( \frac{\partial S_k^i}{\partial x^j} \right) = S_k^j \left( \frac{\partial F_0^i}{\partial x^j} \right) + \Psi_k^j(x) \quad (35)$$

with the  $\Psi_k^j$  known functions of  $x$  (as they depend on the known  $F_k$  and on the  $S_j$  with  $j < k$ , determined at previous stages).

Notice also that  $F_0$  is just the (known) linear part of  $X$ . If we write it in matrix form as  $F_0^i(x) = A_{ij}x^j$  (we will write similarly  $S_0^i(x) = B_{ij}x^j$  for  $S_0$ ), then (35) reads simply

$$[A_{ij}x^j] \left( \frac{\partial S_k^i}{\partial x^j} \right) = A_{ij}S_k^j + \Psi_k^j(x). \quad (36)$$

Solving the determining equations in such a recursive way only requires one to solve at each stage a system of (inhomogeneous) linear PDEs for the  $S_k$ .

For further reference, we introduce the notation  $X_M$  for the linear vector field associated with the matrix  $M$ , i.e.  $X_M = (M_{ij}x^j)(\partial/\partial x^i)$ . We also write  $L_M$  for the homological operator associated with a matrix  $M$ ,  $L_M(.) = \{Mx, .\}$  (see also note 4). In this notation, for any two matrices  $A, B$  we have  $[X_A, X_B] = -X_{[A, B]}$ ; similarly we have, as a consequence of the Jacobi identity:

**Lemma 2** For any matrices  $A, B$ , the commutator of the associated homological operators is given by  $[L_A, L_B] = L_{[B, A]} = -L_{[A, B]}$ .

Let us follow explicitly the iterative procedure for solving (32, 33) for the first steps. For  $k = 0$ , we require that  $\{F_0, S_0\} = 0$ . With our matrix notation,

$$\{F_0, S_0\} \equiv \{Ax, Bx\} = -[A, B]_{ij}x^j \quad (37)$$

and therefore at this stage we only have to determine the matrices  $B$  commuting with a given matrix  $A$ .

For  $k = 1$ , we just get  $\{F_0, S_1\} + \{F_1, S_0\} = 0$ ; in the matrix notation this just reads  $\{Ax, S_1\} = \{Bx, F_1\}$  or, using the homological operators notation,

$$\mathcal{A}(S_1) = \mathcal{B}(F_1) := \Psi_1(x). \quad (38)$$

For  $k = 2$  we get in the same way

$$\mathcal{A}(S_2) = \mathcal{B}(F_2) - \{F_1, S_1\} := \Psi_2(x), \quad (39)$$

and so on for any  $k$ .

**Remark 3** The fact that we can proceed recursively (and only deal with linear PDEs) does not mean that we are guaranteed to find solutions at any stage  $k$ . At  $k = 0$  we always have at least the solutions given by  $B = I$  and by  $B = A^q$  ( $q = 1, 2, \dots$ ). For  $k \geq 1$  we do not, in general, have solutions to the determining equations apart from  $S_j = cF_j$  for all  $q = 0, \dots, k$  (i.e., as  $k$  is generic,  $Y = cX$ ). This corresponds to the fact that symmetry is not generic.

**Remark 4** As the relevant equations are not homogeneous,  $S_k = 0$  is not, in general, an acceptable solution at stage  $k$ . This is quite natural if one thinks that the choice  $B = I$  is always acceptable at  $k = 0$ . Choosing  $S_k = 0$  at all the following stages would leave us with the dilation vector field  $Y = x^i \partial_i$ , which is a symmetry only for  $X$  linear (Gaeta 1994; Olver 1986; Stephani 1989).

### Approximate Symmetries

Note that it could happen where we are only able to determine a commuting vector field for  $X$  up to some finite order  $k$  (either for a full symmetry does not exist or for our limited capacities, computational or otherwise).

If in this case we consider a neighborhood of the origin of small size, say a ball  $\mathcal{B}_\varepsilon$  of radius  $\varepsilon \ll 1$ , in this we have  $[X, Y] = O(\varepsilon^k)$ ; thus,  $Y$  represents an approximate symmetry for  $X$ .

Approximate symmetries are interesting and useful in a number of contexts. In particular, in some cases – notably, for Hamiltonian vector fields – there is a connection between symmetries of dynamical systems and conserved quantities (i.e. constants of motion) for it; in this case, approximate symmetries will correspond to approximate constants of motion, i.e. to quantities which are not exactly conserved, but are approximately so.

More precisely, an approximate symmetry will correspond to a quantity  $J$  whose evolution under the dynamics described by  $X$  is slow of order  $k$ , i.e.  $dJ/dt \approx \varepsilon^k$  for some finite  $k$ . It is rather clear that these can be quite useful in applications, where we are often concerned with study of the dynamics over finite times; see (Cicogna and Gaeta 1999) and especially, in the Hamiltonian case, (Giorgilli 1988).

## Symmetry Characterization of Normal Forms

Let us now consider the case where the dynamical system (13) has already been taken into Poincaré–Dulac normal form. We start by recalling some notions of linear algebra of use here.

### Linear Algebra

A real matrix  $T$  is *semisimple* if its complexification  $T_C$  can be diagonalized, and is *normal* if it commutes with its adjoint,  $[T, T^+] = 0$ . A diagonal matrix is normal. For a normal matrix,  $T : \text{Ker}(T^+) \rightarrow \text{Ker}(T^+)$ . If  $T$  is normal we actually have  $\text{Ker}(T) = \text{Ker}(T^+)$ .

Any semisimple matrix can be transformed into a normal one by a linear transformation. If two semisimple matrices  $A, B$  commute, then they can be simultaneously diagonalized (by a linear, in general non-orthogonal, transformation), and so taken simultaneously to be normal. Thus, when considering such a pair of matrices, we can with no loss of generality assume them to be diagonal or, a fortiori, normal.

If we want to transform  $T$  into a real normal matrix, we just have to consider the transformation of  $T$  into a block diagonal matrix, the blocks corresponding to (complex conjugate) eigenvalues. It is easy to see that in this way we still get a (real) normal matrix. <sup>(8)</sup>

In the following, we will at several points restrict, for ease of discussion, to normal matrices; our statements for normal matrices will be easily extended to semisimple ones up to the appropriate linear transformation.

### Normal Forms

We now note that  $F \in \text{Ker}(\mathcal{A})$  means that the vector field associated with  $F$  (which we denote by  $X_F$ ) commutes with the linear vector field  $X_A$  associated with  $A$ . That is,

$$\begin{aligned} F \in \text{Ker}(\mathcal{A}) &\Leftrightarrow \{Ax, F(x)\} = 0 \\ &\Leftrightarrow [X_A, X_F] = 0. \end{aligned} \quad (40)$$

Thus we have the following characterization for vector fields in normal form (note this uses

Lemma 1 and hence the scalar product defined in section “[Perturbation Theory: Normal Forms](#)”.

**Lemma 3** A vector field  $X = (A_{ij}x^j + F^i(x))\partial_i$ , where the  $F$  are nonlinear functions, is in normal form if and only if its nonlinear part  $X_F = F^i(x)\partial_i$  commutes with the vector field  $X_{A^+} = [A_{ij}^+x^j]\partial_i$  associated with the adjoint of its linear part, i.e. if and only if  $F \in \text{Ker}(\mathcal{A}^+)$ .

For the sake of simplicity, we will only consider the case where the matrix  $A$ , corresponding to the linear part of  $X$  (in both the original and the normal form coordinates), commutes with its adjoint, i.e. we make the following<sup>(9)</sup>

**Assumption** The matrix  $A$  is normal:  $[A, A^+] = 0$ .

If  $A$  is normal, then it follows  $[L_A, L_{A^+}] = 0$  due to Lemma 2; this implies in particular that:

**Lemma 4** If  $A$  is normal, then  $\text{Ker}(\mathcal{A}) = \text{Ker}(\mathcal{A}^+)$ .

It is important to recall that with the standard scalar product, we have  $(L_A)^+ = L_{A^+}$ . It is also important, although trivial, to note that if  $A$  is normal, then  $\mathcal{A}$  is a normal operator (under the standard scalar product), and  $\text{Ker}(\mathcal{A}) \cap \text{Ran}(\mathcal{A}) = \{0\}$ , but  $f \in \text{Ker}(\mathcal{A}^2) \Rightarrow (\mathcal{A}(f)) \in \text{Ker}(\mathcal{A})$  and therefore  $\mathcal{A}(f) = 0$ . Hence

**Lemma 5** If  $A$  is normal,  $\text{Ker}(L_A^2) = \text{Ker}(L_A)$ .

This discussion leads to a natural characterization of Poincaré–Dulac normal forms in terms of symmetry properties.<sup>(10)</sup>

**Lemma 6** If  $A$  is a normal matrix, then  $X = (Ax + F)^i(\partial/\partial x^i)$  is in normal form if and only if  $X_A$  is a symmetry of  $X$ , i.e.  $F \in \text{Ker}(\mathcal{A})$ .

### The General Case

In the general case, i.e. when  $A$  is not normal, the resonant terms will be those in  $\text{Ker}(\mathcal{A}^+)$ ;

similarly, in this case we could characterize systems in normal form by  $\{F(x), A^+x\} = 0$ .

However, for a symmetry characterization it is better to proceed in a slightly different way. That is, we recall that any matrix  $A$  can be uniquely decomposed as  $A = A_s + A_n$  where  $A_s$  is semi-simple and  $A_n$  is nilpotent, with  $[A_s, A_n] = 0$ . Resonance properties involve eigenvalues of  $A_s$ , and resonant terms will satisfy  $\{F(x), A_s^+x\} = 0$ .

This is a more convenient characterization, in that it shows that the full vector field (in normal form)  $X$  will commute with the linear vector field  $S = (A_s)^i x^j \partial_i$  corresponding to the semisimple part of its linear part.

### Symmetries and Transformation to Normal Form

We want to consider the case where the dynamical system (1), or equivalently the vectorfield  $X$ , admits a symmetry  $Y$  (the case of an  $n$ -dimensional symmetry algebra will be considered later on); we want to discuss how the presence of the symmetry affects the normalization procedure. Moreover, as the dynamical and symmetry vector fields are on equal footing, it will be natural to investigate if they can be both put into normal form; or even if some kind of joint normal form is possible (as is the case).

We will use the notation introduced above for the expression of  $X, Y$  in the  $x$  coordinates, and denote by  $y$  the normal form coordinates. Correspondingly, the bracket  $\{.,.\}$  (which is defined in coordinates) will be denoted as  $\{.,.\}_{(x)}$  or  $\{.,.\}_{(y)}$  when confusion is possible. We have, therefore,

$$\begin{aligned} X &= f^i(x)(\partial/\partial x^i) = g^i(y)(\partial/\partial y^i) \\ Y &= s^i(x)(\partial/\partial x^i) = r^i(y)(\partial/\partial y^i) \end{aligned} \quad (41)$$

and similarly for the power series expansions of  $f, g, s, r$  in terms homogeneous of degree  $(k+1) = 1, 2, \dots$ . We will denote the matrices associated with the linear parts of  $X, Y$  by, respectively,  $A$  and  $B : (Df)(0) = (Dg)(0) = A, (Ds)(0) = (Dr)(0) = B$ . The corresponding homological operators will be

denoted by  $\mathcal{A} = L_A$  and by  $\mathcal{B} = L_B$ . We assume that both  $A$  and  $B$  are normal matrices.

### Nonlinear Symmetries (The General Case)

The key, albeit trivial, observation is that the geometric relation  $[X, Y] = 0$  does not depend on the coordinate system we are using. Therefore, if  $\{f, s\}_{(x)} = 0$ , we must also have  $\{g, r\}_{(y)} = 0$ .

Another important, and again quite trivial, observation is that when we consider a P-transformation  $x = y + h_k(y)$ , the term  $F_k$  of order  $k$  in  $X$  changes according to  $\mathcal{A}(h_k)$ , but the term  $S_k$  of order  $k$  in  $Y$  changes according to  $\mathcal{B}(h_k)$ . The same applies when we consider a Lie-Poincaré transformation.

Thus, although when we choose  $h_k + \ell_k$  [with  $\ell_k \in \text{Ker}(\mathcal{A})$ ] to generate the Poincaré transformation, we get the same transformation as that generated by  $h_k$  on the  $F_k$  (see Remark 1), the transformation on  $S_k$  can be different. This means that the freedom left by the Poincaré prescription for construction of the normalizing transformation could, in principle, be used to take the symmetry vector field  $Y$  into some convenient form. This is indeed the case, as will be shown below.

Two vector fields  $X, Y$  on  $M$ , as in (35), with  $A$  and  $B$  semisimple, are in **Joint Normal Form** if both  $G_k$  and  $R_k$  are in  $\text{Ker}(\mathcal{A}) \cap \text{Ker}(\mathcal{B})$  for all  $k \geq 1$ .

**Theorem 3** Let the vector fields  $X = f^i(x)\partial_i$  and  $Y = s^i(x)\partial_i$  have a fixed point in  $x_0 = 0$ . Let them commute,  $[X, Y] = 0$ , and have normal semisimple linear parts  $A = (Df)(0)$  and  $B = (Ds)(0)$ . Then, by means of a sequence of Poincaré transformations, they can be brought to Joint Normal Form.

In this theorem (a proof of this is given in (Cicogna and Gaeta 1994b, 1999); see also (Gramchev and Yoshino 1999) for a different approach to a related problem) we did not really use the interpretation of one of the vector fields as describing the dynamics of the system and the other describing a symmetry, but only their commutation relation. From this point of view, it is

natural to also consider arbitrary (possibly non-Abelian) algebras of vector fields.

### Linear Symmetries

A special case of symmetries is given by *linear* symmetries, i.e. by the case where

$$Y = X_B = (B_{ij}x^j)(\partial/\partial x^i). \quad (42)$$

In this case, if  $X, Y$  are in Joint Normal Form we have in particular that  $f \in \text{Ker}(\mathcal{B})$ . We have the following corollaries to Theorem 3:

**Corollary 1** If the linear vector field  $Y = X_B$  is a symmetry for  $X = f^i(x)\partial_i$ , it is possible to normalize  $f$  by passing to  $y$  coordinates so that  $X = g^i(y)(\partial/\partial y^i)$ ,  $Y = (By)^i(\partial/\partial y^i)$ , and  $g \in \text{Ker}(\mathcal{A}) \cap \text{Ker}(\mathcal{B})$ .

**Corollary 2** Let  $s(x) = Bx + S(x)$ , with  $S$  the nonlinear part of  $s$ , and let  $Y = s^i(x)\partial_i$  be a symmetry of  $X = f^i(x)\partial_i$ . Then when  $X, Y$  are put in Joint Normal Form,  $X_B$  is a linear symmetry of  $X$ .

When we perform the Poincaré transformations needed to transform  $X$  into its normal form, there seems to be no reason, a priori, why the  $Y$  should keep its linear form. It is actually possible to prove the following result (the proof of this relies on a similar result by Ruelle (1973, 1989) dealing with the center manifold mapping, and is given e.g. in (Elphick et al. 1987; Iooss and Adelmeyer 1992); see also (Belitskii 1978, 1981, 1987, 2002)) that we quote from (Iooss and Adelmeyer 1992):

**Theorem 4** If  $X$  commutes with a linear vector field  $Y = X_B = (B_{ij}x^j)\partial_i$ , then it is possible to find a normalizing series of Poincaré transformations with generators  $h_k \in \text{Ker}(\mathcal{B})$ , so that in the new coordinates  $y$ ,  $X$  is taken into normal form and  $Y$  is left unchanged, i.e.  $Y = (B_{ij}y^j)\partial_i$ .

Note that, for resonant  $B$ , Theorem 4 is *not* a special case of Theorem 3 and the above Corollaries. <sup>(11)</sup>



**Remark 5** One should avoid confusions between linear symmetries of the dynamical system and symmetries of its linearization, which do not extend in general to symmetries of the full system.

## Generalizations

In this section we are going to discuss some generalizations of the results illustrated in the previous one: we will deal with the problem of transformation into normal form of a (possibly non Abelian) Lie algebra with more than two generators, not necessarily commuting.<sup>(12)</sup>

### Abelian Lie Algebra

It is actually convenient to drop the distinction between the vector field defining the dynamical system and the symmetry vector fields. Thus, we simply consider an algebra  $\mathcal{G}$  of vector fields  $X_i$ .<sup>(13)</sup>

First of all, we consider the case of an *Abelian* Lie algebra of vector fields. In this case we have the following result (see Cicogna and Gaeta (1999) for a proof).

**Theorem 5** Let  $\{X_1, \dots, X_r\}$  commute, and assume that the matrices  $A_{(i)}$  identifying the linear parts of the  $X_j$  are normal. Then  $\{X_1, \dots, X_r\}$  can be put in Joint Normal Form by a sequence of Poincaré or Lie-Poincaré transformations.

Note that if  $\mathcal{A}_i$  are the homological operators corresponding to the  $A_{(i)}$ , and we write  $\mathcal{K} = \cap_{i=0}^k \text{Ker}(\mathcal{A}_i)$ , the Theorem states there are coordinates  $y^i$  such that  $X_j = \left( \tilde{f}_{(j)}(y) \right)_i (\partial/\partial y^i)$  with  $\tilde{f}_{(j)} \in \mathcal{K}$  for all  $j$ .

### Nilpotent Lie Algebra

For generic Lie algebras one cannot expect results as general as in the Abelian case (Arnal et al. 1984; Cicogna and Gaeta 1994a, 1999). A significant exception to this is met in the case of *nilpotent* algebras<sup>(14)</sup> (see also Cushman and Sanders (1986) for a group-theoretical approach), in which we can recover essentially the same results obtained in the Abelian case (see Hermann (1968) for the case of semisimple Lie algebras).

Actually, an extension of Theorem 5 to the nilpotent case should be considered with some care, as the only nilpotent algebras of nontrivial semisimple matrices are Abelian. On the other hand, we could have a non-Abelian nilpotent algebra of vector fields with linear parts given by semisimple matrices, provided that some of these vanish.

Indeed, although  $[X_i, X_j] = c_{ij}^k X_k$  necessarily implies that  $[A_i, A_j] = c_{ij}^k A_k$ , it could happen that all the  $A_k$  for which there exists a nonzero  $c_{ij}^k$  do vanish, so that the algebra of vector fields is “Abelian at the linear level”. (As a concrete example, consider the algebra spanned by  $X = -x^2 d/dx$  and  $Y = x d/dx$ .)

Note that in this case  $\text{Ker}(\mathcal{A}_k)$  would be just the whole space; needless to say, we should consider the full vector fields  $X_k$ , which will produce (by assumption) a closed Lie algebra.

With this remark (and in view of the fact that the proof of Theorem 5 is based on properties of the algebra of the  $A_i$ 's and of the corresponding homological operators, see Cicogna and Gaeta (1999)) it is to be expected that the result for the nilpotent case will not substantially differ from the one holding for the Abelian case (as usual, the key to this extension will be to proceed to normalization of vector fields in an order which respects the structure of the Lie algebra). This is indeed what happens, and one has the following result (Cicogna and Gaeta 1994a, 1999):

**Theorem 6** Let the vector fields  $\{X_1, \dots, X_r\}$  form a nilpotent Lie algebra  $\mathcal{G}$  under  $[\cdot, \cdot]$ ; assume that the matrices  $A_{(i)}$  identifying the linear parts of the  $X_j$  are normal. Then  $\{X_1, \dots, X_r\}$  can be put in Joint Normal Form by a sequence of Poincaré or Lie-Poincaré transformations.

**Corollary 3** If the general Lie algebra  $\mathcal{G}$  of vector fields  $\{X_1, \dots, X_n\}$  contains a nilpotent subalgebra  $\mathcal{G}^*$ , then the set of vector fields  $X_i$  spanning this  $\mathcal{G}^*$  can be put in Joint Normal Form.

### General Lie Algebra

Some “partial” Joint Normal Form can be obtained, even for non-nilpotent algebras, under some special assumptions. We will just quote a



result in this direction, referring as usual to (Cicogna and Gaeta 1999) for details. A description of normal forms for systems with symmetry corresponding to simple compact Lie groups is given in (Gaeta 2002b).

**Theorem 7** Let  $\mathcal{G}$  be a  $d$ -dimensional algebra spanned by  $X_{F_a}$  with  $F_a = A_a x + F_a$  ( $a = 1, \dots, d$ ), and let  $\mathcal{G}$  admit a non-trivial center  $C(\mathcal{G})$ . Let the center of  $\mathcal{G}$  be spanned by  $X_{w_b}$  with  $w_b = C_b x + W_b$  ( $b = 1, \dots, d_C$ , where  $d_C \leq d$ ), and assume that the semisimple parts  $C_{b,s}$  are normal matrices. Denote by  $C_{b,s}$  the associated homological operators, and write  $\mathcal{K}_s = \cap_{b=1}^{d_C} \text{Ker}(C_{b,s})$ .

Then, by means of a sequence of Poincaré transformations, all the  $F_a$  can be taken into  $\hat{F}_a \in \mathcal{K}_s$ . The same holds for  $\hat{\mathcal{G}} = \mathcal{G} \oplus \mathcal{N}$ , with  $\mathcal{N}$  a nilpotent subalgebra.

## Symmetry for Systems in Normal Form

No definite relation exists, in general, between symmetries  $\mathcal{G}_X$  of a vector field  $X$  and symmetries  $\mathcal{G}_A$  of its linear part, or between constants of motion  $\mathcal{I}_X$  for  $X$  and constants of motion  $\mathcal{I}_A$  for its linear part, but if  $X$  is in normal form, one has some interesting results (Cicogna and Gaeta 1994a; Walcher 1991):

**Lemma 7** If  $X$  is in normal form, any constant of motion of  $X$  must also be a constant of motion of its linearization  $A$ , i.e.  $\mathcal{I}_X \subseteq \mathcal{I}_A$ .

In general  $\mathcal{I}_X \neq \mathcal{I}_A$ , even if  $X$  is in normal form. Also, if  $[B, A] = 0$  in general  $\mu(x)Bx$  does not belong to  $\mathcal{G}_X$ , even for  $\mu \in \mathcal{I}_X$  (unless  $X_B \in \mathcal{L}_X$  as well, where  $\mathcal{L}_X$  are the linear symmetries of  $X$ ), nor to  $\mathcal{G}_A$  (unless  $\mu \in \mathcal{I}_A$ ).

**Lemma 8** If  $X$  is in normal form, then  $\mathcal{G}_X \subseteq \mathcal{G}_A$ .

This result allows for the restricting our search for  $Y \in \mathcal{G}_X$  to  $\mathcal{G}_A$  rather than considering the full set of vector fields on  $M$ . Similarly, Lemmas 7 and 8 can be useful in the determination of the sets  $\mathcal{L}_X$ ,  $\mathcal{I}_X$ , in that we can first solve the easier problem of

determining  $\mathcal{L}_A$ ,  $\mathcal{I}_A$ , and then look for  $\mathcal{L}_X$ ,  $\mathcal{I}_X$  in the class of vector fields  $\mathcal{L}_A$  and of the functions in  $\mathcal{I}_A$ , rather than in the full set of linear vector fields on  $M$  and, respectively, in the full set of scalar functions on  $M$ .

Moreover,  $\mathcal{G}_A$  can be determined in a relatively simple way, by solving the system of quasi-linear non-homogeneous first order PDEs  $\{Ax, g\} = 0$ , which are written explicitly as

$$(A_{ij}x^j)(\partial g^k / \partial x^i) = A_{kj}g^j. \quad (43)$$

By considering this, and introducing the set  $\mathcal{I}_A^*$  of the meromorphic (i.e., quotients of formal power series) constants of motion of the linear problem  $\dot{x} = Ax$ , one can obtain (Cicogna and Gaeta 1994a; Elphick et al. 1987; Walcher 1991) the following result:

**Lemma 9**  $\mathcal{G}_A$  is the set of all formal power series in  $\mathcal{I}_A^* \otimes \mathcal{L}_A$ .

In a more explicit form, as  $\mathcal{G}_A \equiv \text{Ker}(\mathcal{A})$ , the resonant terms  $F(x) \in \text{Ker}(\mathcal{A})$  are power series of the form  $F(x) = K(\rho(x)) \cdot x$ , where  $K$  is a matrix commuting with  $A$  when written in terms of its real entries  $K_{ij}$ , and where  $K(\rho(x))$  is the same matrix in which the entries  $K_{ij}$  are replaced by functions of the constants of motion  $\rho = \rho(x) \in \mathcal{I}_A^*$ . The set of the vector fields in  $\mathcal{G}_A$  is of course a Lie algebra.

We summarize our discussion for dynamical systems in normal form in the following proposition:

**Theorem 8** Let  $X$  be a vector field in normal form, and let  $A$  be the normal matrix corresponding to its linear part. Then,  $\mathcal{L}_X \subseteq \mathcal{G}_X \subseteq \mathcal{G}_A$ ; and  $\mathcal{L}_X \subseteq \mathcal{L}_A \subseteq \mathcal{G}_A$ .

**Remark 6** It should be mentioned that Kodama considered the problem of determining  $\mathcal{G}_A$  from a more algebraic standpoint (Kodama 1994). In the same work, Kodama also observed that  $\mathcal{G}_A$ , considered as an algebra, is not only infinite-dimensional, but has the natural structure of a graded Virasoro algebra.

**Remark 7** We emphasize once again that the above results were given for  $X$  in normal form. They can obviously be no longer true if  $X$  is not in normal form. <sup>(15)</sup>

## Linearization of a Dynamical System

An interesting application of the Joint Normal Form deals with the case of linearizable dynamical systems. Clearly, if  $\text{Ker}(\mathcal{A}) = \{0\}$ , the dynamical system is linearizable by means of a formal Poincaré transformation. But, whatever the matrix  $A$ , the linear vector field  $X_A = (Ax \cdot \nabla)$  commutes with the vector field  $S = \sum x_i (\partial/\partial x_i) = ((Lx) \cdot \nabla)$ , which generates the dilations in  $\mathbf{R}^n$ . It is easy to see that, conversely, the only vector fields commuting with  $S$  are the linear ones. It is also clear that the identity does not admit resonances. Thus (Bambusi et al. 1998; Gaeta and Marmo 1996):

**Lemma 10** A vector field  $X_f$  (or a dynamical system  $\dot{x} = f(x)$ ) can be linearized if and only if it admits a (possibly formal) symmetry  $X_g$  such that  $B = (Dg)(0) = I$ .

Proceeding in a similar way – but using Joint Normal Forms – we have, more generally:

**Theorem 9** The vector field  $X_f$  with  $A = (Df)(0)$  can be linearized if and only if it admits a (possibly formal) symmetry  $X_g$  with  $B = (Dg)(0)$  such that  $A$  and  $B$  do not admit common resonances, i.e.  $\text{Ker}(\mathcal{A}) \cap \text{Ker}(\mathcal{B}) = \{0\}$ .

This result can be easily extended not only to the case of more than one symmetry, as an obvious consequence of Theorem 4, but also to the non-semisimple case (Cicogna and Gaeta 1994a).

Another interesting result related to linear dynamical systems, is the following (Cicogna and Gaeta 1994a, 1999):

**Theorem 10** If a dynamical system in  $\mathbf{R}^n$  can be linearized, then it admits  $n$  independent *commuting* symmetries, that can be simultaneously linearized.

## Further Normalization and Symmetry

As repeatedly noted above (see in particular Remark 1), when the linear part of the dynamical vector field is resonant the resulting degeneracy in the solution to the homological equation is not a real degeneracy for what concerns effects on higher order terms in the normal form.

These higher order terms could – and in general will – generate resonant terms, which cannot be eliminated by the standard algorithm. On the other hand, it is clear that this could be seen as a bonus rather than as a problem: in fact, it is conceivable that by carefully choosing the component of  $h_k$  lying in  $\text{Ker}(\mathcal{A})$ , one could generate resonant terms which exactly cancel those already present in the vector field.

Several algorithms have been designed to take advantage, in one way or the other, of this possibility; some review of different approaches is provided in the paper (Chen and Della Dora 2000). Here we are concerned with those based on symmetry properties, and discuss two different approaches, developed respectively by the present author (Gaeta 1997, 1999b) and by Palacian and Yanguas (2001).

It should be stressed that once the presence of additional symmetries – and in the Hamiltonian context, additional constants of motion – has been determined for the normal form truncated at some finite order  $N$ , one should investigate if the set of symmetries (or constants of motion) persist under small perturbations; in particular, when considering terms of higher order as well. A general tool to investigate this kind of question is provided by the Nekhoroshev generalization of the Poincaré–Lyapounov theorem (Nekhoroshev 1994, 2002, 2005); see also the discussion in (Bambusi and Gaeta 2002; Gaeta 2002c, 2003, 2006c).

## Further Normalization and Resonant Further Symmetry

We will assume again, for ease of discussion, that the matrix  $A$  associated with the linear part of the dynamical vector field  $X$  is normal. In this case, as discussed above, the normal form is written as  $X = g^i(x) \partial_i$  where  $g(x) = \sum_{k=0}^{\infty} G_k(x)$ , with

$G_k \in \mathcal{V}_k$  and all  $G_k$  being resonant. We can correspondingly write

$$X = \sum_{k=0}^{\infty} X_k, X_k = G_k^i \partial_i. \quad (44)$$

As  $X$  is in normal forms, we are guaranteed to have

$$[X_0, X_k] = 0 \quad \forall k = 0, 1, 2, \dots; \quad (45)$$

this corresponds to the characterization of normal forms in terms of symmetry as discussed in section “[Symmetry Characterization of Normal Forms](#)”. In other words,  $G_k \in \text{Ker}(\mathcal{A})$  and hence, defining  $\mathcal{G}_0$  as the Lie algebra of vector fields commuting with  $X_0$ ,  $X_k \in \mathcal{G}_0$ . On the other hand, in general it will be  $[X_j, X_k] \neq 0$  for  $j \neq 0$  and both  $j, k$  greater than zero; thus  $\mathcal{G}_0$  is not, in general, an Abelian Lie algebra: we can only state that  $X_0$  belongs to the center of  $\mathcal{G}_0$ ,  $X_0 \in Z(\mathcal{G}_0)$ .

Suppose we want to operate further Lie-Poincaré transformations generated by functions which are symmetric under  $X_0$  (i.e. are in the kernel of  $\mathcal{A}$ ). It follows from the formulas obtained in section “[Perturbation Theory: Normal Forms](#)” that these will map  $\mathcal{G}_0$  into itself, i.e.  $\mathcal{G}_0$  is globally invariant under this restricted set of Lie-Poincaré transformations.

As for the individual vector fields, it follows from the general formula (24) that each of them is invariant under such a transformation *at first order* in  $h_k$ , but not at higher orders. That is, making use again of Remark 1, we can still in this way generate new resonant terms in the normal form, including maybe terms which cancel some of those present in (44).

By looking at the explicit formulas (25), it is rather easy to analyze in detail the higher order terms generated in the concerned Lie-Poincaré transformation.

Note that we did not take into account problems connected with the convergence of the further normalizing transformations; it has to be expected that each step will reduce the radius of convergence, so that the further normalized forms will be actually (and not just formally) conjugated

to the original dynamic in smaller and smaller neighborhoods of the fixed point; we refer to (Gaeta 2006b) for an illustration of this point by explicit examples and numerical computations; and to ▶ “[Convergence of Perturbative Expansions](#)” for a general discussion on the convergence of normalizing transformations.

In order to state exactly the result obtained by this construction, we need to introduce some function spaces, which require abstract definitions. With <sup>(16)</sup>  $\mathcal{L}_k(\cdot) := [X_k, \cdot]$ , we set

$$H^{(p)} := \text{Ker}(\mathcal{L}_0) \cap \dots \cap \text{Ker}(\mathcal{L}_{p-1}) \quad (46)$$

(note  $H^{(q)} \subseteq H^{(p)}$  for  $q > p$ ), and denote by  $\mathcal{M}_p$  the restriction of  $\mathcal{L}_p$  to  $H^{(p)}$ ; thus  $\text{Ker}(\mathcal{M}_p) = H^{(p+1)}$ . We define spaces  $F^{(p)}$  (with  $F^{(0)} = \mathcal{V}$ ) as

$$F^{(p)} := \text{Ker}(\mathcal{M}_0^+) \cap \dots \cap \text{Ker}(\mathcal{M}_{p-1}^+); \quad (47)$$

the adjoint should be meant in the sense of the scalar product introduced in section “[Perturbation Theory: Normal Forms](#)”. We also write  $F_k^{(p)} = F^{(p)} \cap \mathcal{V}_k$ .

We will say that  $X$  is in Poincaré renormalized form up to order  $N$  if  $G_k \in F_k^{(k)}$  for all  $k \leq N$ .

**Theorem 11** The vector field  $X$  can be formally taken into Poincaré renormalized form up to (any finite) order  $N$  by means of a (finite) sequence of Lie-Poincaré transformations.

For a proof of this theorem, and a detailed description of the renormalizing procedure, see (Cicogna and Gaeta 1999; Gaeta 1997, 1999b, 2002a); see (Cicogna and Gaeta 1999; Gaeta 1997) for the case where additional symmetries are present; an improved procedure, taking full advantage of the Lie algebraic structure of  $\mathcal{G}_0$ , is described in (Gaeta 2001).

### Further Normalization and External Symmetry

A different approach to further reduction (simplification) of vector fields in normal form has been developed by Palacian and Yanguas (2001) (and applied mainly in the context of Hamiltonian dynamics (Palacián and Yanguas 2000,

2003, 2005)), making use of a result by Meyer (Meyer and Hall 1992).

As discussed in the previous subsection we have  $[X_0, X_k] = 0$  for all  $k$ . Suppose now there is a linear (in the coordinates used for the decomposition (44)) vector field  $Y$  such that  $[X_0, Y] = 0$ . Then the Jacobi identity guarantees that

$$[X_0, [X_k, Y]] = 0 \quad \forall k. \quad (48)$$

We assume that  $Y$  also corresponds to a normal matrix  $B$ , so that the homological operator  $\mathcal{B}$  associated to it is also normal.

We can then proceed to further normalization as above, being guaranteed that – provided we choose  $h_k \in \text{Ker}(\mathcal{L}_0)$  – the resulting vector fields will not only still be in  $\mathcal{G}_0$ , but also still satisfy (48). One can use freedom in the choice of the generator  $h_k \in \text{Ker}(\mathcal{L}_0)$  for the further normalization in a different way than discussed above: that is, we can choose it so that  $[Y, X_k] = 0$ ; in other words, we will get  $\hat{X} \in \text{Ker}(\mathcal{A}) \cap \text{Ker}(\mathcal{B})$ .

Note the advantage of this: we do not have to worry about complicated matters related to relevant homological operators acting between different spaces, as we only make use of the homological operator  $\mathcal{B}$  associated with the “external” symmetry linear vector field and thus mapping each  $\mathcal{V}_k$  into the same  $\mathcal{V}_k$ .

The result one can obtain in this way is the following (which we quote in a simplified setting for the sake of brevity; in particular the normality assumption can be relaxed) (Palacián and Yanguas 2001).

**Theorem 12** Let  $X$  be in normal form; assume moreover  $X_0 = Ax$  and there is a normal matrix  $B$  such that  $[A, B] = 0$ ; denote by  $Y$  the associated vector field,  $Y = (Bx)^i \partial_i$ . Assume moreover for each resonant vector field  $R_k$  there is  $Q_k$  satisfying  $[Y, Q_k] = 0$  and such that  $\hat{R}_k = R_k + [X_0, Q_k]$  commutes with  $Y$ . Then  $X$  can be taken to a (different) normal form  $\hat{X}$  such that  $[Y, \hat{X}] = 0$ .

Applications of this theorem, and more generally of this approach, are discussed e.g. in (Palacián and Yanguas 2000, 2001, 2003, 2005). We stress that albeit the assumptions of this

theorem are rather strong<sup>(17)</sup>, it points out to the fact that *there are* cases in which a symmetry – and in the Hamiltonian framework, an integral of motion – of the linear part can be extended to a symmetry of the full normal form.

## Symmetry Reduction of Symmetric Normal Forms

Symmetry reduction is a general – and powerful – approach to the study of nonlinear dynamical systems. (In the Hamiltonian case, this is also known as (Marsden–Weinstein) moment map (Arnold 1976; Marsden 1992; Marsden and Ratiu 1994).)

A general theory based on the geometry of group action has been developed by Louis Michel; this was originally motivated by the study of spontaneous symmetry breaking in high-energy physics (Michel and Radicati 1971, 1973) (see (Michel 1971a, b, 1975, 1980) for the simpler case where only stationary solutions are considered, and (Michel and Zhilinskii 2001) for the full theory and applications; see also (Abud and Sartori 1983; Gaeta 1999a, 2002a, 2006a; Gaeta and Morando 1997; Sartori 1991, 2002)).

A description of this would lead us too far away from the scope of this paper, but as this theory also applies to vector fields in normal form, we will briefly describe the results that can be obtained in this way; we will mainly follow (Gaeta and Walcher 2005) (see (Gaeta and Walcher 2006; Gaeta et al. 2008) for further detail, extensions, and a more abstract mathematical formulation).

As mentioned in section “Symmetry of Dynamical Systems”, the Lie algebra of vector fields in normal form is infinite dimensional, but also has the structure of a Lie module over the algebra of constants of motion for the linear part  $X_0$  of the vector field (which remains the same under all the considered transformations).

Let us recall that vector monomial  $\mathbf{v}_{\mu, \alpha} := x^\mu \mathbf{e}_\alpha$  is *resonant* with  $A$  if

$$\begin{aligned} (\mu \cdot \lambda) &:= \sum_{i=1}^n \mu_i \lambda_i = \lambda_\alpha \quad \text{with } \mu_i \geq 0 \\ |\mu| &:= \sum_{i=1}^n \mu_i \geq 1 \end{aligned} \quad (49)$$

here  $\lambda_i$  are the eigenvalues of  $A$ , which we suppose to be semisimple, for the sake of simplicity (in the general case one would consider  $A_s$  rather than  $A$ ).

As mentioned in section “[Perturbation Theory: Normal Forms](#)”, the relation  $(\mu \cdot \lambda) = \lambda_\alpha$  is said to be a resonance relation related to the eigenvalue  $\lambda_\alpha$ , and the integer  $|\mu|$  is said to be the *order* of the resonance. In the present context it is useful to include order one resonances in the definition (albeit the *trivial* order one resonances given by  $\lambda_\alpha = \lambda_\alpha$  are obviously of little interest).

Let us consider again the resonance Eq. (49). It is clear that if there are non-negative integers  $\sigma_i$  (some of them nonzero) such that

$$\sum_{i=1}^n \sigma_i \lambda_i = 0, \quad (50)$$

then we always have infinitely many resonances. In this case the monomial  $\varphi = x^\sigma$  will be called a resonant scalar monomial. It is an invariant of  $X_0$ , and any multi-index  $\mu$  with  $\mu_i = k\sigma_i + \delta_{i\alpha}$  provides a resonance relation  $(\mu \cdot \lambda) = \lambda_\alpha$  related to the eigenvalue  $\lambda_\alpha$ ; in other words, any monomial  $x^{k\sigma} x^\alpha = \varphi^k x^\alpha$  is resonant, and so is any vector  $\mathbf{v}_{k\sigma + e_\alpha, \alpha}$ .

Therefore, we say that (50) identifies a invariance relation. The presence of invariance relations is the only way to have infinitely many resonances in a finite dimensional system (see (Walcher 1991)).

Any nontrivial resonance (49) which does not originate in an invariance relation, is said to be a sporadic resonance. Sporadic resonances are always in finite number (if any) in a finite dimensional system (Walcher 1991).

Any invariance relation (50) such that there is no  $v$  with  $v_i \leq \sigma_i$  (and of course  $v \neq \sigma$ ) providing another invariance relation, is said to be an elementary invariance relation. Every invariance relation is a linear combination (with nonnegative integer coefficients) of elementary ones. Elementary invariance relations are always in finite number (if any) in a finite dimensional system (Walcher 1991). If there are  $m$  independent elementary invariance relations, each of them of the

form (50), we associate to these monomials  $\beta_j = x^\sigma = \prod_{i=1}^n x_i^{\sigma_i}$  ( $j = 1, \dots, m$ ).

Similarly, if there are  $r$  sporadic resonances (49), we associate resonant monomials  $\alpha^j(x) = x^\mu = \prod_{i=1}^n x_i^{\mu_i}$  ( $j = 1, \dots, r$ ) and resonant vectors  $\mathbf{v}_{\mu, \alpha}^{(j)}$  to sporadic resonances. We then introduce two set of new coordinates: these will be the coordinates  $w^1, \dots, w^r$  in correspondence with sporadic resonances, and other new coordinates  $\varphi^1, \dots, \varphi^m$  in correspondence with elementary invariance relations. The evolution equations for the  $x^i$  can be written in simplified form using these (note that some ambiguity is present here, in that we can write these in different ways in terms of the  $x, w, \varphi$ ), but we should also assign evolution equations for them; these will be given in agreement with the dynamics itself. That is, we set

$$\frac{dw^j}{dt} = \frac{\partial w^j}{\partial x^i} \frac{dx^i}{dt} := h^j(x, w, \varphi); \quad (51)$$

$$\frac{d\varphi^a}{dt} = \frac{\partial \varphi^a}{\partial x^i} \frac{dx^i}{dt} := z^a(x, w, \varphi). \quad (52)$$

We are thus led to consider the enlarged space  $W = \mathbf{R}^{n+r+m}$  of the  $(x, w, \varphi)$ , and in this the vector field

$$Y = f^i(x, w, \varphi) (\partial / \partial x^i) + h^j(x, w, \varphi) (\partial / \partial w^j) + z^a(x, w, \varphi) (\partial / \partial \varphi^a) \quad (53)$$

The vector field  $Y$  is uniquely defined on the manifold identified by  $\psi^j := w^j - \alpha^j(x) = 0$ ,  $\varphi^a - \beta^a(x) = 0$ . It is obvious (by construction) that the  $(n+m)$ -dimensional manifold  $M \subset W$  identified by  $\psi^j := w^j - \alpha^j(x) = 0$  is invariant under the flow of  $Y$ , see (51). It is also easy to show that the functions  $z^a$  defined in (52) can be written in terms of the  $f$  variables alone, i.e.  $\partial z^a / \partial x^i = \partial z^a / \partial w^j = 0$ . This implies<sup>(18)</sup>

**Lemma 11** The evolution of the  $\varphi$  variables is described by a (generally, nonlinear) equation involving the  $\varphi$  variables alone.



Note that the equations for  $x$  and  $w$  depend on  $f$  and are therefore non-autonomous. We have the following result (we refer to (Gaeta and Walcher 2005, 2006) for a proof; see (Gaeta et al. 2008) for extensions).

**Theorem 13** The analytic functions  $f^i$  and  $h^i$  defined above can be written as linear in the  $x$  and  $w$  variables, the coefficients being functions of the  $\varphi$  variables. Hence the evolution of the  $x$  and  $w$  variables is described by non-autonomous linear equations, obtained by inserting the solution  $\varphi = \varphi(t)$  of the equations for  $\varphi$  in the equations  $\dot{x} = f(x, w, \varphi)$ ,  $\dot{w} = h(x, w, \varphi)$ .

Note that if no invariance relations are present, hence no  $\varphi$  variables are introduced, then the system describing the time evolution of the  $x$ ,  $w$  variables is linear; in this case we can interpret normal forms as projections of a linear system to an invariant manifold (without symmetry reduction).

If there are no sporadic resonances of order greater than one, then upon solving the reduced equation for the  $\varphi$  variables one obtains a non-autonomous linear system. Moreover, if all eigenvalues are distinct then we have a product system of one-dimensional equations.

If  $\varphi(t)$  converges to some constant  $\varphi_0$ , the asymptotic evolution of the system is governed by a linear autonomous equation for  $x$  and  $w$ . Similarly, if there is a periodic solution  $\overline{\varphi}(t)$  and  $\varphi(t)$  converges to  $\overline{\varphi}(t)$  for large  $t$ , the asymptotic evolution of the system is governed by a linear equation with  $t$ -periodic coefficients for  $x$  and  $w$ .

## Conclusions

We have reviewed the basic notions concerning symmetry of dynamical systems and its determination, in particular in a perturbative setting. We have subsequently considered various situations where the interplay of perturbation theory and symmetry properties produce nontrivial results, either in that the perturbative expansion turns out to be simplified (with respect to the general case) due to symmetry properties, or in that

computations of symmetry is simplified by dealing with a system in normal form; we then considered the problem of jointly normalizing an algebra of vector fields (with possibly but not necessarily one of these defining a dynamical system, the other being symmetry vector fields). We also discussed how normal forms can be characterized in terms of symmetry, and how this is extended to a characterization of “renormalized forms”. Finally we considered symmetry reduction applied to systems in normal form.

The discussion conducted here illustrates some of the powerful conceptual and computational simplifications arising for systems with symmetry, also in the realm of perturbation theory.

As remarked in the Introduction, symmetry is a non-generic property; on the other hand, it is often the case that equations arising from physical (mechanical, electronic, etc.) systems enjoy some degree of symmetry, as a consequence of the symmetry of the fundamental equations of physics. Disregarding the symmetry properties in these cases would mean renouncing the use of what is often the only handle to grab the behavior of non-linear systems; and correspondingly a symmetry analysis can often on the one hand lead to identifying several relevant properties of the system even without a complete solution, and on the other hand being instrumental in obtaining exact (or approximate, as we are here dealing with perturbation theory) solutions.

Here we discussed some of the consequences of symmetry for the specific case of dynamical systems, such as those met in analyzing the behavior of nonlinear systems near a known (e.g. trivial) solution.

For a more general discussion, as well as for concrete applications, the reader is referred on the one hand to texts discussing symmetry for differential equations (Aleekseevskij et al. 1991; Baumann 2000; Bluman and Anco 2002; Bluman and Kumei 1989; Cantwell 2002; Gaeta 1994; Hydon 2000; Krasil’shchik and Vinogradov 1999; Olver 1986, 1995; Stephani 1989), on the other to texts and papers specifically dealing with the interplay of symmetry and perturbation theory, quoted in the main text and listed in the ample Bibliography below.

## Future Developments

First and foremost, future developments should be expected to concern further application of the general theory in concrete cases, both in nonlinear theoretical mechanics and in specific subfields, ranging from the more applied (e.g., ship dynamics and stability (Bakri et al. 2004; Tondl et al. 2000); or handling of complex electrical networks (Lin et al. 1996; Vittal et al. 1998)) to the more theoretical (e.g., galactic dynamics (Belmonte et al. 2006; de Zeeuw and Merritt 1983)).

On the other hand, the theory developed so far is in many cases purely formal, in that consideration of convergence properties – and estimation of the convergence region in phase and/or parameter space – of the resulting (perturbative) series is often left aside, with the understanding that in any concrete application one will have explicit series and be ready to analyze their convergence. The story of general (i.e. non-symmetric) perturbation theory shows however that the theoretical analysis of convergence properties can be precious – not only for the conceptual understanding but also in view of concrete applications – and it should thus be expected that future developments will deal with convergence properties in the symmetric case (see e.g. ► “Convergence of Perturbative Expansions”, (Cicogna and Walcher 2002) for a review of existing results).

The same issue of convergence, and estimation of the convergence region, arises in connection with further normalization (under any approach), and has so far been given little consideration.

A different kind of generalization is called for when dealing with symmetry reduction of normal forms: in fact, on the one hand it is natural to try applying the same approach (based on quite general geometrical properties) to more general systems than initially considered; on the other hand the method discussed in section “Symmetry Reduction of Symmetric Normal Forms” is algorithmic and could be implemented by symbolic manipulations packages – such as those already existing for computations of symmetry of differential equations – which would be of help to anybody having to deal with perturbation of concrete symmetric systems.

Finally, another field of future developments can be called for: here we discussed the interplay of perturbation theory and “standard” symmetries. Or, the notion of symmetry of differential equations has been generalized in various directions, producing in some cases a significant advantage in application to concrete systems. It should thus be expected that the interplay between these generalized notions of symmetry and perturbation theory will be investigated in the near future, and most probably will produce interesting and readily applicable results.

## Additional Notes

We collect here the additional notes called throughout the manuscript.

1. Note that  $\tilde{M}$  is naturally a fiber bundle (Chern et al. 1999; Isham 1999; Nakahara 1990; Nash and Sen 1983) over  $t$ : that is, it can be decomposed as the union of copies of  $M$ , each one in correspondence to a value of  $t$ . A *section* of this bundle is simply the graph of a function  $\sigma : t \rightarrow M$ . The set  $\sigma_x$  considered in a moment is a section of this fiber bundle.
2. For general differential equations, one would go along the same lines. A relevant difference is however present: for (systems of) first order equations there is no algorithmic way to find the general solutions to determining equations, as opposed to any other case (Olver 1986; Stephani 1989).
3. In the case of Hamiltonian systems, one can work directly on the Hamiltonian  $H(p, q)$  rather than on the Hamilton equations of motion ► “Hamiltonian Perturbation Theory (and Transition to Chaos)”, (Hanssmann 2007). We will however, for the sake of brevity, not discuss the specific features of the Hamiltonian case.
4. Later on, in particular when we deal with different homological operators, it will be convenient to also denote this  $\mathcal{A}$  as  $L_A$ , with reference to the matrix  $A$  appearing in  $F_0 = Ax$ .
5. This is equivalent to defining  $(\xi_{\mu, i}, \xi_{\nu, j}) = \delta_{\mu, \nu} \delta_{i, j} (\mu!)$ , where for the multi-index  $\mu$  we have defined  $\mu! = (\mu_1!) \dots (\mu_n!)$ .



6. The name “resonant” is due to the relation existing between eigenvectors of  $\mathcal{A}$  and resonance relations among eigenvalues of the matrix  $A = (Df)(0)$  describing the linear part  $F_0(x) = Ax$  of the system.
7. In order to determine  $\tilde{X} = \tilde{f}^i(x)\partial/\partial x^i$ , we write (using the exponential notation)  $\xi(t + \delta t) = e^{\delta t X} \xi(t)$ . Therefore,  $x(t + \delta t) = [e^{sH} e^{(\delta t) X} \xi(t)]_s = 1$ . Using  $\xi(t) = e^{-H} x(t)$ , we have  $[x(t + \delta t) - x(t)] = [e^{sH}(e^{(\delta t)X} - I)e^{-sH}x(t)]_{(s=1)}$ , and therefore (24).
8. Note, however, that when the diagonalizing matrix  $M$  is not unitary, this transformation changes implicitly the scalar product.
9. The general case is discussed e.g. in (Cicogna and Gaeta 1999); see also, for a more general discussion on normal forms with non-normal and non-semisimple normal form, the article ► “Perturbation of Systems with Nilpotent Real Part”.
10. Note that the terms in  $\text{Ker}(\mathcal{A}) = \text{Ker}(\mathcal{A}^+)$  are just the resonant ones. Thus, the present characterization of normal forms is equivalent to the one given earlier on.
11. We also note that in the framework of Lie–Poincaré transformations, i.e. when considering the time one action of a vectorfield  $F$  (Benettin et al. 1984; Bogoliubov and Mitropolsky 1961; Mitropolsky and Lopatin 1995) (see above), Theorem 4 shows that  $F$  can be chosen to admit  $Y$  as a symmetry; see (Cicogna and Gaeta 1999).
12. Generalizations can also be obtained in the direction of relaxing the normality assumption for the matrices identifying the linear part of vector fields. We will not discuss this case, referring the reader instead to ► “Perturbation of Systems with Nilpotent Real Part”, (Cicogna and Gaeta 1999).
13. If this is the symmetry algebra of one of the vector fields, say  $X_0$ , we have that  $[X_0, X_i] = 0$  for all the  $i$ , i.e.  $X_0$  belongs to the center of the algebra  $\mathcal{G}$ .
14. We stress that we refer here to the case of a nilpotent Lie algebra, not to the case where the relevant matrices are nilpotent!
15. This is readily shown by the following example. Consider the two-dimensional system, which is *not* in normal form,  $\dot{x}_1 = -x_1 x_2, \dot{x}_2 = x_2$ . It is easily seen that  $\mu = x_1 \exp(x_2)$  is a constant of motion, but not of the linearized problem, i.e.  $\mu \in \mathcal{I}_X$ , but  $\mu \notin \mathcal{I}_A$ . Similarly, we have that  $Y = x_1^2 \exp(x_2)(\partial/\partial x_1) \in \mathcal{G}_X$  but  $Y \notin \mathcal{G}_A$ .
16. It should be noted that  $X_k$ , and hence  $\mathcal{L}_k$ , change under further normalization transformations; however, they stabilize after a finite number of steps, and in particular will not change anymore after the further normalization reaches their order.
17. Note that the  $\widehat{X}_k$  in Theorem 12 satisfies  $\mathcal{B}(\widehat{X}_k) := [Y, \widehat{X}_k] = [Y, X_k] + [Y, [X_0, Q_k]]$ ; using the Jacobi identity, and assuming  $[Y, Q_k] = 0$ , this reads  $\mathcal{B}(X_k) - \mathcal{A}[\mathcal{B}(Q_k)]$ . On the other hand,  $[A, B] = 0$  guarantees  $[\mathcal{A}, \mathcal{B}] = 0$ , hence we get  $\mathcal{B}(\widehat{X}_k) = \mathcal{B}[X_k - \mathcal{A}(Q_k)]$ . Thus the assumption of Theorem 12 could be rephrased in terms of kernels of the operators  $\mathcal{A}^+$  and  $\mathcal{B}$  as follows: for each  $X_k \in \text{Ker}(\mathcal{A}^+) \cap \mathcal{V}_k$ , there exists  $Q_k \in \text{Ker}(\mathcal{B})$  such that  $[X_k - \mathcal{A}(Q_k)] \in \text{Ker}(\mathcal{B})$ .
18. As the  $\varphi$  identify group orbits for the group  $G$  generated by the Lie algebra, we interpret  $\dot{\varphi} = z(\varphi)$  as an equation in orbit space, and the equation for  $(x, w)$  as an equation on the Lie group  $G$ . Methods for the solution of the latter are discussed in (Wei and Norman 1963), see also (Carinena et al. 2000).

## Bibliography

- Abenda S, Gaeta G, Walcher S (eds) (2003) Symmetry and perturbation theory – SPT2002. In: Proceedings of Cala Gonone workshop, 19–26 May 2002. World Scientific, Singapore
- Abud M, Sartori G (1983) The geometry of spontaneous symmetry breaking. *Ann Phys* 150:307–372
- Alekseevskij DV, Vinogradov AM, Lychagin VV (1991) Basic ideas and concepts of differential geometry. In: Gamkrelidze RV (ed) *Encyclopaedia of mathematical sciences vol 28 – Geometry I*. Springer, Berlin
- Arnald D, Ben Ammar M, Pinczon G (1984) The Poincaré–Dulac theorem for nonlinear representations of nilpotent Lie algebras. *Lett Math Phys* 8:467–476
- Arnold VI (1974) *Equations différentielles ordinaires*. MIR, Moscow, 2nd edn 1990. Arnold VI (1992) *Ordinary differential equations*. Springer, Berlin

- Arnold V (1976) Les méthodes mathématiques de la mécanique classique. MIR, Moscow. Arnold VI (1983, 1989) Mathematical methods of classical mechanics. Springer, Berlin
- Arnold V (1980) Chapitres supplémentaires de la théorie des équations différentielles ordinaires. MIR, Moscow. Arnold VI (1983) Geometrical methods in the theory of ordinary differential equations. Springer, Berlin
- Arnold VI, Il'yashenko YS (1988) Ordinary differential equations. In: Anosov DV, Arnold VI (eds) Encyclopaedia of mathematical sciences vol 1 – Dynamical systems I. Springer, Berlin, pp 1–148
- Arnold VI, Kozlov VV, Neishtadt AI (1993) Mathematical aspects of classical and celestial mechanics. In: Arnold VI (ed) Encyclopaedia of mathematical sciences vol 3 – Dynamical systems III, 2nd edn. Springer, Berlin, pp 1–291
- Baider A (1989) Unique normal form for vector fields and Hamiltonians. *J Diff Eqs* 78:33–52
- Baider A, Churchill RC (1988) Uniqueness and non-uniqueness of normal forms for vector fields. *Proc R Soc Edinb A* 108:27–33
- Baider A, Sanders J (1992) Further reduction of the Takens-Bogdanov normal form. *J Diff Eqs* 99:205–244
- Bakri T, Nabergoj R, Tondl A, Verhulst F (2004) Parametric excitation in non-linear dynamics. *Int J Nonlinear Mech* 39:311–329
- Bambusi D, Gaeta G (eds) (1997) Symmetry and perturbation theory. In: Proceedings of Torino workshop, ISI, December 1996. GNFM–CNR, Roma
- Bambusi D, Gaeta G (2002) On persistence of invariant tori and a theorem by Nekhoroshev. *Math Phys Electron J* 8:1–13
- Bambusi D, Cicogna G, Gaeta G, Marmo G (1998) Normal forms, symmetry, and linearization of dynamical systems. *J Phys A Math Gen* 31:5065–5082
- Bambusi D, Gaeta G, Cadoni M (2001) Symmetry and perturbation theory – SPT 2001. In: Proceedings of the international conference SPT 2001, Cala Gonone, 6–13 May 2001. World Scientific, Singapore
- Bargmann V (1961) On a Hilbert space of analytic functions and an associated integral transform. *Commun Pure Appl Math* 14:187–214
- Baumann G (2000) Symmetry analysis of differential equations with mathematica. Springer, New York
- Belitskii GR (1978) Equivalence and normal forms of germs of smooth mappings. *Russ Math Surv* 33(1):107–177
- Belitskii GR (1981) Normal forms relative to the filtering action of a group. *Trans Moscow Math Soc* 40(2): 1–39
- Belitskii GR (1987) Smooth equivalence of germs of vector fields with a single eigenvalue or a pair of purely imaginary eigen values. *Funct Anal Appl* 20:253–259
- Belitskii GR (2002)  $C^\infty$ -Normal forms of local vector fields. *Acta Appl Math* 70:23–41
- Belmonte C, Boccaletti D, Pucacco G (2006) Stability of axial orbits in galactic potentials. *Celest Mech Dyn Astron* 95:101–116
- Benettin G, Galgani L, Giorgilli A (1984) A proof of the Kolmogorov theorem on invariant tori using canonical transformations defined by the Lie method. *Nuovo Cimento B* 79:201–223
- Bluman GW, Anco SC (2002) Symmetry and integration methods for differential equations. Springer, Berlin
- Bluman GW, Kumei S (1989) Symmetries and differential equations. Springer, Berlin
- Bogoliubov NN, Mitropolsky VA (1961) Asymptotic methods in the theory of nonlinear oscillations. Hindustan, New Delhi. (1962) Méthodes asymptotiques dans la théorie des oscillations non-linéaires. Gauthier-Villars, Paris
- Broer HW (1979) Bifurcations of singularities in volume preserving vector fields. Ph.D Thesis, Groningen
- Broer HW (1981) Formal normal form theorems for vector fields and some consequences for bifurcations in the volume preserving case. In: Rand DA, Young LS (eds) Dynamical systems and turbulence. IT Lect Notes Math 898. Springer, Berlin
- Broer HW, Takens F (1989) Formally symmetric normal forms and genericity. *Dyn Rep* 2:39–59
- Bryuno AD (1971a) Analytical form of differential equations I. *Trans Moscow Math Soc* 25:131–288
- Bryuno AD (1971b) Analytical form of differential equations II. *Trans Moscow Math Soc* 26:199–239
- Bryuno AD (1988) The normal form of a Hamiltonian system. *Russ Math Surv* 43(1):25–66
- Bryuno AD (1989) Local methods in the theory of differential equations. Springer, Berlin
- Bryuno AD, Walcher S (1994) Symmetries and convergence of normalizing transformations. *J Math Anal Appl* 183:571–576
- Cantwell BJ (2002) Introduction to symmetry analysis. Cambridge University Press, Cambridge
- Carinena JF, Grabowski J, Marmo G (2000) Lie-Scheffers systems: a geometric approach. Bibliopolis, Napoli
- Chen G, Della Dora J (2000) Further reductions of normal forms for dynamical systems. *J Diff Eqs* 166: 79–106
- Chern SS, Chen WH, Lam KS (1999) Lectures on differential geometry. World Scientific, Singapore
- Chossat P (2002) The reduction of equivariant dynamics to the orbit space for compact group actions. *Acta Appl Math* 70:71–94
- Chossat P, Lauterbach R (1999) Methods in equivariant bifurcations and dynamical systems with applications. World Scientific, Singapore
- Chow SN, Hale JK (1982) Methods of bifurcation theory. Springer, Berlin
- Chow SN, Li C, Wang D (1994) Normal forms and bifurcations of planar vector fields. Cambridge University Press, Cambridge
- Chua LO, Kokubu H (1988) Normal forms for nonlinear vector fields Part I: theory. *IEEE Trans Circ Syst* 35: 863–880
- Chua LO, Kokubu H (1989) Normal forms for nonlinear vector fields Part II: applications. *IEEE Trans Circ Syst* 36:851–870
- Churchill RC, Kummer M, Rod DL (1983) On averaging, reduction and symmetry in Hamiltonian systems. *J Diff Eqs* 49:359–414

- Cicogna G, Gaeta G (1994a) Normal forms and nonlinear symmetries. *J Phys A* 27:7115–7124
- Cicogna G, Gaeta G (1994b) Poincaré normal forms and Lie point symmetries. *J Phys A* 27:461–476
- Cicogna G, Gaeta G (1994c) Symmetry invariance and center manifolds in dynamical systems. *Nuovo Cim B* 109:59–76
- Cicogna G, Gaeta G (1999) *Symmetry and perturbation theory in nonlinear dynamics*. Springer, Berlin
- Cicogna G, Walcher S (2002) Convergence of normal form transformations: the role of symmetries. *Acta Appl Math* 70:95–111
- Courant R, Hilbert D (1962) *Methods of mathematical physics*. Wiley, New York; (1989)
- Crawford JD (1991) Introduction to bifurcation theory. *Rev Mod Phys* 63:991–1037
- Crawford JD, Knobloch E (1991) Symmetry and symmetry-breaking bifurcations in fluid dynamics. *Ann Rev Fluid Mech* 23:341–387
- Cushman R, Sanders JA (1986) Nilpotent normal forms and representation theory of  $s\mathfrak{l}_2$ ,  $R$ . In: Golubitsky M, Guckenheimer J (eds) *Multi-parameter bifurcation theory*. Contemp Math 56, AMS, Providence
- de Zeeuw T, Merritt D (1983) Stellar orbits in a triaxial galaxy I Orbits in the plane of rotation. *Astrophys J* 267:571–595
- Degasperis A, Gaeta G (eds) (1999) *Symmetry and perturbation theory II – SPT98*. In: Proceedings of Roma workshop, Università La Sapienza, December 1998. World Scientific, Singapore
- Deprit A (1969) Canonical transformation depending on a small parameter. *Celest Mech* 1:12–30
- Elphick C, Tirapegui E, Brachet ME, Coulet P, Iooss G (1987) A simple global characterization for normal forms of singular vector fields. *Physica D* 29:95–127. (1988) Addendum. *Physica D* 32:488
- Fassò F (1990) Lie series method for vector fields and Hamiltonian perturbation theory. *ZAMP* 41:843–864
- Fassò F, Guzzo M, Benettin G (1998) Nekhoroshev stability of elliptic equilibria of Hamiltonian systems. *Commun Math Phys* 197:347–360
- Field MJ (1989) *Equivariant bifurcation theory and symmetry breaking*. *J Dyn Diff Eqs* 1:369–421
- Field MJ (1996a) *Lectures on bifurcations, dynamics and symmetry*. Res Notes Math 356. Pitman, Boston
- Field MJ (1996b) Symmetry breaking for compact Lie groups. *Mem AMS* 574:1–170
- Field MJ, Richardson RW (1989) Symmetry breaking and the maximal isotropy subgroup conjecture for reflection groups. *Arch Rat Mech Anal* 105:61–94
- Field MJ, Richardson RW (1990) Symmetry breaking in equivariant bifurcation problems. *Bull Am Math Soc* 22:79–84
- Field MJ, Richardson RW (1992a) Symmetry breaking and branching patterns in equivariant bifurcation theory I. *Arch Rat Mech Anal* 118:297–348
- Field MJ, Richardson RW (1992b) Symmetry breaking and branching patterns in equivariant bifurcation theory II. *Arch Rat Mech Anal* 120:147–190
- Fokas AS (1979a) Generalized symmetries and constants of motion of evolution equations. *Lett Math Phys* 3: 467–473
- Fokas AS (1979b) Group theoretical aspects of constants of motion and separable solutions in classical mechanics. *J Math Anal Appl* 68:347–370
- Fokas AS (1980) A symmetry approach to exactly solvable evolution equations. *J Math Phys* 21: 1318–1326
- Fokas AS (1987) Symmetries and integrability. *Stud Appl Math* 77:253–299
- Fokas AS, Gelfand IM (1996) Surfaces on Lie groups, Lie algebras, and the integrability. *Commun Math Phys* 177:203–220
- Fontich E, Gelfreich VG (1997) On analytical properties of normal forms. *Nonlinearity* 10:467–477
- Forest E, Murray D (1994) Freedom in minimal normal forms. *Physica D* 74:181–196
- Fushchich WI, Nikitin AG (1987) *Symmetries of Maxwell equations*. Reidel, Dordrecht
- Fushchich WI, Shtelen WM, Slavutsky SL (1989) *Symmetry analysis and exact solutions of nonlinear equations of mathematical physics*. Naukova Dumka, Kiev
- Gaeta G (1990) Bifurcation and symmetry breaking. *Phys Rep* 189:1–87
- Gaeta G (1994) Nonlinear symmetries and nonlinear equations. Kluwer, Dordrecht
- Gaeta G (1997) Reduction of Poincaré normal forms. *Lett Math Phys* 42:103–114 & 235
- Gaeta G (1999a) An equivariant branching lemma for relative equilibria. *Nuovo Cim B* 114:973–982
- Gaeta G (1999b) Poincaré renormalized forms. *Ann IHP Phys Theor* 70:461–514
- Gaeta G (2001) Algorithmic reduction of Poincaré–Dulac normal forms and Lie algebraic structure. *Lett Math Phys* 57:41–60
- Gaeta G (2002a) Poincaré normal and renormalized forms. *Acta Appl Math* 70:113–131
- Gaeta G (2002b) Poincaré normal forms and simple compact Lie groups. *Int J Mod Phys A* 17:3571–3587
- Gaeta G (2002c) The Poincaré–Lyapounov–Nekhoroshev theorem. *Ann Phys* 297:157–173
- Gaeta G (2003) The Poincaré–Nekhoroshev map. *J Nonlinear Math Phys* 10:51–64
- Gaeta G (2006a) Finite group symmetry breaking. In: Francoise JP, Naber G, Tsou ST (eds) *Encyclopedia of mathematical physics*. Kluwer, Dordrecht
- Gaeta G (2006b) Non-quadratic additional conserved quantities in Birkhoff normal forms. *Celest Mech Dyn Astron* 96:63–81
- Gaeta G (2006c) The Poincaré–Lyapounov–Nekhoroshev theorem for involutory systems of vector fields. *Ann Phys N Y* 321:1277–1295
- Gaeta G, Marmo G (1996) Nonperturbative linearization of dynamical systems. *J Phys A* 29:5035–5048
- Gaeta G, Morando P (1997) Michel theory of symmetry breaking and gauge theories. *Ann Phys N Y* 260: 149–170

- Gaeta G, Walcher S (2005) Dimension increase and splitting for Poincaré-Dulac normal forms. *J Nonlinear Math Phys* 12:S1327–S1342
- Gaeta G, Walcher S (2006) Embedding and splitting ordinary differential equations in normal form. *J Diff Eqs* 224:98–119
- Gaeta G, Prinari B, Rauch S, Terracini S (eds) (2005) Symmetry and perturbation theory – SPT2004. In: Proceedings of Cala Gonone workshop, 30 May – 6 June 2004. World Scientific, Singapore
- Gaeta G, Vitolo R, Walcher S (eds) (2007) Symmetry and perturbation theory – SPT2007. In: Proceedings of Otranto workshop, 2–9 June 2007. World Scientific, Singapore
- Gaeta G, Grosshans FD, Scheurle J, Walcher S (2008) Reduction and reconstruction for symmetric ordinary differential equations. *J Diff Eqs* 244:1810–1839
- Gallavotti G (1983) The elements of mechanics. Springer, Berlin
- Giorgilli A (1988) Rigorous results on the power expansions for the integrals of a Hamiltonian system near an elliptic equilibrium point. *Ann IHP Phys Theor* 48: 423–439
- Giorgilli A, Locatelli U (1997) Kolmogorov theorem and classical perturbation theory. *ZAMP* 48:220–261
- Giorgilli A, Morbidelli A (1997) Invariant KAM tori and global stability for Hamiltonian systems. *ZAMP* 48: 102–134
- Giorgilli A, Zehnder E (1992) Exponential stability for time dependent potentials. *ZAMP* 43:827–855
- Glendinning P (1994) Stability, instability and chaos: an introduction to the theory of nonlinear differential equations. Cambridge University Press, Cambridge
- Golubitsky M, Stewart I, Schaeffer D (1988) Singularity and groups in bifurcation theory, vol II. Springer, Berlin
- Gramchev T, Yoshino M (1999) Rapidly convergent iteration methods for simultaneous normal forms of commuting maps. *Math Z* 231:745–770
- Guckenheimer J, Holmes P (1983) Nonlinear oscillations, dynamical systems, and bifurcation of vector fields. Springer, Berlin
- Gustavson FG (1964) On constructing formal integrals of a Hamiltonian system near an equilibrium point. *Astron J* 71:670–686
- Guzzo M, Fassò F, Benettin G (1998) On the stability of elliptic equilibria. *Math Phys Electron J* 4(1):16
- Hamermesh M (1962) Group theory. Addison-Wesley, Reading; reprinted by Dover, New York (1991)
- Hanssmann H (2007) Local and semi-local bifurcations in Hamiltonian dynamical systems results and examples. Springer, Berlin
- Hermann R (1968) The formal linearization of a semi-simple Lie algebra of vector fields about a singular point. *Trans AMS* 130:105–109
- Hoveijn I (1996) Versal deformations and normal forms for reversible and Hamiltonian linear systems. *J Diff Eqs* 126:408–442
- Hoveijn I, Verhulst F (1990) Chaos in the 1:2:3 Hamiltonian normal form. *Physica D* 44:397–406
- Hydon PE (2000) Symmetry methods for differential equations. Cambridge UP, Cambridge
- Ibragimov N (1992) Group analysis of ordinary differential equations and the invariance principle in mathematical physics. *Russ Math Surv* 47(4):89–156
- Il'yashenko YS, Yakovenko SY (1991) Finitely smooth normal forms of local families of diffeomorphisms and vector fields. *Russ Math Surv* 46(1):1–43
- Iooss G, Adelmeyer M (1992) Topics in bifurcation theory and applications. World Scientific, Singapore
- Isham CJ (1999) Modern differential geometry for physicists. World Scientific, Singapore
- Kinyon M, Walcher S (1997) On ordinary differential equations admitting a finite linear group of symmetries. *J Math Anal Appl* 216:180–196
- Kirillov AA (1976, 1984) Elements of the theory of representations. Springer, Berlin
- Kodama Y (1994) Normal forms, symmetry and infinite dimensional Lie algebra for systems of ODE's. *Phys Lett A* 191:223–228
- Kokubu H, Oka H, Wang D (1996) Linear grading function and further reduction of normal forms. *J Diff Eqs* 132: 293–318
- Krasil'shchik IS, Vinogradov AM (1984) Nonlocal symmetries and the theory of coverings. *Acta Appl Math* 2: 79–96
- Krasil'shchik IS, Vinogradov AM (1999) Symmetries and conservation laws for differential equations of mathematical physics. AMS, Providence
- Kummer M (1971) How to avoid secular terms in classical and quantum mechanics. *Nuovo Cimento B* 1: 123–148
- Kummer M (1976) On resonant nonlinearly coupled oscillators with two equal frequencies. *Commun Math Phys* 48:53–79
- Lamb J (1996) Local bifurcations in  $k$ -symmetric dynamical systems. *Nonlinearity* 9:537–557
- Lamb J (1998)  $k$ -symmetry and return maps of spacetime symmetric flows. *Nonlinearity* 11:601–630
- Lamb J, Melbourne I (2007) Normal form theory for relative equilibria and relative periodic solutions. *Trans AMS* 359:4537–4556
- Lamb J, Roberts J (1998) Time reversal symmetry in dynamical systems: a survey. *Physica D* 112:1–39
- Levi D, Winternitz P (1989) Non-classical symmetry reduction: example of the Boussinesq equation. *J Phys A* 22:2915–2924
- Lin CM, Vittal V, Kliemann W, Fouad AA (1996) Investigation of modal interaction and its effect on control performance in stressed power systems using normal forms of vector fields. *IEEE Trans Power Syst* 11: 781–787
- Marsden JE (1992) Lectures on mechanics. Cambridge University Press, Cambridge
- Marsden JE, Ratiu T (1994) Introduction to mechanics and symmetry. Springer, Berlin
- Meyer KR, Hall GR (1992) Introduction to Hamiltonian dynamical systems and the N-body problem. Springer, New York

- Michel L (1971a) Points critiques de fonctions invariantes sur une  $G$ -variété. *Comptes Rendus Acad Sci Paris* 272-A:433–436
- Michel L (1971b) Nonlinear group action Smooth action of compact Lie groupson manifolds. In: Sen RN, Weil C (eds) *Statistical mechanics and field theory*. Israel University Press, Jerusalem
- Michel L (1975) Les brisure spontanées de symétrie en physique. *J Phys Paris* 36-C7:41–51
- Michel L (1980) Symmetry defects and broken symmetry configurations hidden symmetry. *Rev Mod Phys* 52: 617–651
- Michel L, Radicati L (1971) Properties of the breaking of hadronic internal symmetry. *Ann Phys N Y* 66: 758–783
- Michel L, Radicati L (1973) The geometry of the octet. *Ann IHP* 18:185–214
- Michel L, Zhilinskii BI (2001) Symmetry, invariants, topology basic tools. *Phys Rep* 341:11–84
- Mikhailov AV, Shabat AB, Yamilov RI (1987) The symmetry approach to the classification of non-linear equations complete list of integrable systems. *Russ Math Surv* 42(4):1–63
- Mitropolsky YA, Lopatin AK (1995) *Nonlinear mechanics, groups and symmetry*. Kluwer, Dordrecht
- Nakahara M (1990) *Geometry, topology and physics*. IOP, Bristol
- Nash C, Sen S (1983) *Topology and geometry for physicists*. Academic, London
- Nekhoroshev NN (1994) The Poincaré–Lyapunov–Liouville–Arnol’d theorem. *Funct Anal Appl* 28: 128–129
- Nekhoroshev NN (2002) Generalizations of Gordon theorem. *Regul Chaotic Dyn* 7:239–247
- Nekhoroshev NN (2005) Types of integrability on a submanifold and generalizations of Gordons theorem. *Trans Moscow Math Soc* 66:169–241
- Olver PJ (1986) *Applications of Lie groups to differential equations*. Springer, Berlin
- Olver PJ (1995) *Equivalence, invariants, and symmetry*. Cambridge University Press, Cambridge
- Ovsjiannikov LV (1982) *Group analysis of differential equations*. Academic, London
- Palacián J, Yanguas P (2000) Reduction of polynomial Hamiltonians by the construction of formal integrals. *Nonlinearity* 13:1021–1054
- Palacián J, Yanguas P (2001) Generalized normal forms for polynomial vector fields. *J Math Pures Appl* 80: 445–469
- Palacián J, Yanguas P (2003) Equivariant N-DOF Hamiltonians via generalized normal forms. *Comm Cont Math* 5:449–480
- Palacián J, Yanguas P (2005) A universal procedure for normalizing  $n$ -degree-of-freedom polynomial Hamiltonian systems. *SIAM J Appl Math* 65:1130–1152
- Pucci E, Saccomandi G (1992) On the weak symmetry group of partial differential equations. *J Math Anal Appl* 163:588–598
- Ruelle D (1973) Bifurcation in the presence of a symmetry group. *Arch Rat Mech Anal* 51:136–152
- Ruelle D (1989) *Elements of differentiable dynamics and bifurcation theory*. Academic, London
- Sadovskii DA, Delos JB (1996) Bifurcation of the periodic orbits of Hamiltonian systems – an analysis using normal form theory. *Phys Rev A* 54:2033–2070
- Sanders JA (2003) Normal form theory and spectral sequences. *J Diff Eqs* 192:536–552
- Sanders JA (2005) Normal forms in filtered Lie algebra representations. *Acta Appl Math* 87:165–189
- Sanders JA, Verhulst F (1985) *Averaging methods in nonlinear dynamical systems*. Springer, Berlin
- Sanders JA, Verhulst F, Murdock J (2007) *Averaging methods in nonlinear dynamical systems*. Springer, Berlin
- Sartori G (1991) Geometric invariant theory. A model-independent approach to spontaneous symmetry and/or super symmetry breaking. *Riv N Cim* 14–11: 1–120
- Sartori G (2002) Geometric invariant theory in a model-independent analysis of spontaneous symmetry and super symmetry breaking. *Acta Appl Math* 70:183–207
- Sartori G, Valente G (2005) Constructive axiomatic approach to the determination of the orbit spaces of coregular compact linear groups. *Acta Appl Math* 87: 191–228
- Sattinger DH (1979) *Group theoretic methods in bifurcation theory*. Lecture notes in mathematics 762. Springer, Berlin
- Sattinger DH (1983) *Branching in the presence of symmetry*. SIAM, Philadelphia
- Sattinger DH, Weaver O (1986) *Lie groups and algebras*. Springer, Berlin
- Siegel K, Moser JK (1971) *Lectures on celestial mechanics*. Springer, Berlin; reprinted in *Classics in Mathematics*. Springer, Berlin (1995)
- Sokolov VV (1988) On the symmetries of evolutions equations. *Russ Math Surv* 43(5):165–204
- Stephani H (1989) *Differential equations their solution using symmetries*. Cambridge University Press, Cambridge
- Stewart I (1988) *Bifurcation with symmetry*. In: Bedford T, Swift J (eds) *New directions in dynamical systems*. Cambridge University Press, Cambridge
- Tondl A, Ruijgrok T, Verhulst F, Nabergoj R (2000) *Auto-parametric resonance in mechanical systems*. Cambridge University Press, Cambridge
- Ushiki S (1984) Normal forms for singularities of vector fields. *Jpn J Appl Math* 1:1–34
- Vanderbauwhede A (1982) *Local bifurcation and symmetry*. Pitman, Boston
- Verhulst F (1989) *Nonlinear differential equations and dynamical systems*. Springer, Berlin; (1996)
- Verhulst F (1998) *Symmetry and integrability in Hamiltonian normal form*. In: Bambusi D, Gaeta G (eds) *Symmetry and perturbation theory*. CNR, Roma
- Verhulst F (1999) *On averaging methods for partial differential equations*. In: Degasperis A, Gaeta G (eds)

- Symmetry and perturbation theory II. World Scientific, Singapore
- Vinogradov AM (1984) Local symmetries and conservation laws. *Acta Appl Math* 2:21–78
- Vittal V, Kliemann W, Ni YX, Chapman DG, Silk AD, Sobajic DJ (1998) Determination of generator groupings for an islanding scheme in the Manitoba hydro system using the method of normal forms. *IEEE Trans Power Syst* 13:1346–1351
- Vorob'ev EM (1986) Partial symmetries of systems of differential equations. *Soviet Math Dokl* 33:408–411
- Vorob'ev EM (1991) Reduction and quotient equations for differential equations with symmetries. *Acta Appl Math* 23:1–24
- Walcher S (1991) On differential equations in normal form. *Math Ann* 291:293–314
- Walcher S (1993) On transformation into normal form. *J Math Anal Appl* 180:617–632
- Walcher S (1999) Orbital symmetries of first order ODEs. In: Degasperis A, Gaeta G (eds) *Symmetry and perturbation theory II*. World Scientific, Singapore
- Walcher S (2000) On convergent normal form transformations in the presence of symmetry. *J Math Anal Appl* 244:17–26
- Wei J, Norman E (1963) Lie algebraic solution of linear differential equations. *J Math Phys* 4:575–581
- Winternitz P (1987) What is new in the study of differential equations by group theoretical methods? In: Gilmore R (ed) *Group theoretical methods in physics proceedings of the XV ICGTMP*. World Scientific, Singapore
- Winternitz P (1993) Lie groups and solutions of nonlinear PDEs. In: Ibort LA, Rodríguez MA (eds) *Integrable systems, quantum groups, and quantum field theory NATO ASI 9009*. Kluwer, Dordrecht

# Perturbation of Systems with Nilpotent Real Part

Todor Gramchev  
Dipartimento di Matematica e Informatica,  
Università di Cagliari, Cagliari, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Complex and Real Jordan Canonical Forms  
Nilpotent Perturbation and Formal Normal Forms  
of Vector Fields and Maps Near a Fixed Point  
Loss of Gevrey Regularity in Siegel Domains in  
the Presence of Jordan Blocks  
First-Order Singular Partial Differential  
Equations  
Normal Forms for Real Commuting Vector Fields  
with Linear Parts Admitting Nontrivial Jordan  
Blocks  
Analytic Maps near a Fixed Point in the Presence  
of Jordan Blocks  
Weakly Hyperbolic Systems and Nilpotent  
Perturbations  
Bibliography

## Glossary

**Perturbation** Typically, one starts with an “initial” system  $S_0$ , which is usually simple and/or well understood. We perturb the system by adding a (small) perturbation  $R$  so that the new object becomes  $S_0 + R$ . In our context the typical examples for  $S_0$  will be systems of linear ordinary differential equations with constant coefficients in  $\mathbb{R}^n$  or the associated linear vector fields.

**Nilpotent linear transformation** Let  $A : \mathbb{K}^n \mapsto \mathbb{K}^n$  be a linear map, where  $\mathbb{K} = \mathbb{R}$  or

$\mathbb{K} = \mathbb{C}$ . We call  $A$  nilpotent if there exists a positive integer  $r$  such that the  $r$ th iteration  $A^r$  become the zero map, in short  $A^r = 0$ .

**Gevrey spaces** Let  $\Omega$  be an open domain in  $\mathbb{R}^n$  and let  $\sigma \geq 1$ . The Gevrey space  $G^\sigma(\Omega)$  stands for the set of all functions  $f \in C^\infty(\Omega)$  such that for every compact subset  $K \subset \subset \Omega$  one can find  $C = C_{K,f} > 0$  such that

$$\sup_{x \in K} |\partial_x^\alpha f(x)| \leq C^{|\alpha|+1} \alpha!^\sigma \quad (1)$$

for all  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n$ ,  $\alpha! = \alpha_1! \dots \alpha_n!$ ,  $|\alpha| := \alpha_1 + \dots + \alpha_n$ . If  $\sigma = 1$  we recapture the space of real analytic functions in  $\Omega$  while the scale  $G^\sigma(\Omega)$ ,  $\sigma > 1$ , serves as an intermediate space between the real analytic functions and the set of all  $C^\infty$  functions in  $\Omega$ . By the Stirling formula one may replace  $\alpha!^\sigma$  by  $|\alpha|!^\sigma$ ,  $|\alpha|^{\sigma|\alpha|}$  or  $\Gamma(\sigma|\alpha|)$ , where  $\Gamma(z)$  stands for the Euler Gamma function cf. the book of Rodino (1993) for more details on the Gevrey spaces.

One associates also Gevrey index to formal power series, namely, given a (formal) power series

$$f(x) = \sum_{\alpha} f_{\alpha} x^{\alpha}$$

this is in the formal Gevrey space  $G_f^\sigma(\mathbb{K}^n)$  if there exist  $C > 0$  and  $R > 0$  such that

$$|f_{\alpha}| \leq C^{|\alpha|+1} |\alpha|!^{\sigma-1} \quad (2)$$

for all  $\alpha \in \mathbb{Z}_+^n$ .

In fact, one can find in the literature another definition of the formal Gevrey spaces  $G_f^\tau$  of index  $\tau$ , namely replacing  $\sigma - 1$  by  $\tau$  (see e.g. Ramis (1984)).

## Definition of the Subject

The main goal of this article is to dwell upon the influence of the presence (explicit and/or hidden) of nontrivial real nilpotent perturbations appearing in problems in Dynamical Systems, Partial

Todor Gramchev: deceased.

© Springer-Verlag 2009  
G. Gaeta (ed.), *Perturbation Theory*,  
[https://doi.org/10.1007/978-1-0716-2621-4\\_395](https://doi.org/10.1007/978-1-0716-2621-4_395)

Originally published in  
R. A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*, © Springer-Verlag 2009,  
[https://doi.org/10.1007/978-3-642-27737-5\\_395](https://doi.org/10.1007/978-3-642-27737-5_395)



Differential Equations and Mathematical Physics. Under the term nilpotent perturbation we will mean, broadly speaking, a classical linear algebra type setting: we start with an object (vector field or map near a fixed point, first-order singular partial differential equations, system of evolution partial differential equations) whose “linear part”  $A$  is semi simple (diagonalizable) and we add a (small) nilpotent part  $N$ . The problems of interest might be summarized as follows: are the “relevant properties” (in suitable functional framework) of the initial “object” stable under the perturbation  $N$ . If not, to classify, if possible, the novel features of the perturbed systems.

Broadly speaking, the cases when the instabilities occur are rare, they form some kind of exceptional sets. However, they appear in important problems (both in mathematical and physical contexts) when degeneracies (bifurcations) occur.

## Introduction

We will focus our attention on topics where the presence of nontrivial Jordan blocks in the linear parts changes the properties of the original systems (i.e., instabilities occurs unless additional restrictions are imposed):

- (i) Convergence/divergence issues for the normal form theory of vector fields and maps near a singular (fixed) point in the framework of spaces of analytic functions and Gevrey classes.
- (ii) (Non)solvability for singular partial differential equations near a singular point.
- (iii) Cauchy problems for hyperbolic systems of partial differential equations with multiple characteristics.

Some basic features of the normal form theory for vector fields near a point will be recalled with an emphasis on the difficulties appearing in the presence of nontrivial Jordan blocks in the classification and computational aspects of the normal forms. For more details and various aspects of perturbation theory in Dynamics we refer to other articles in the Perturbation Theory Section of this Encyclopedia:

cf. Bambusi ▶ “[Perturbation Theory for PDEs](#)”, Gaeta ▶ “[Symmetry and Perturbation Theory in Non-linear Dynamics](#)”, Gallavotti ▶ “[Perturbation Theory](#)”, Broer ▶ “[Normal Forms in Perturbation Theory](#)”, Broer and Hansmann ▶ “[Hamiltonian Perturbation Theory \(and Transition to Chaos\)](#)”, Teixeira ▶ “[Perturbation Theory for Non-smooth Systems](#)”, Verhulst ▶ “[Perturbation Analysis of Parametric Resonance](#)”, Walcher ▶ “[Convergence of Perturbative Expansions](#)”.

We start by outlining some motivating examples.

Consider the nilpotent planar linear system of ordinary differential equations

$$\dot{x} = \begin{pmatrix} 0 & \varepsilon \\ 0 & 0 \end{pmatrix} x, \quad x(0) = x^0 \in \mathbb{R}^2,$$

where  $\varepsilon \in \mathbb{R}$ . The explicit solution is given by  $x_1(t) = x_1^0 + x_2^0 \varepsilon t$ ,  $x_2(t) = x_2^0$  and clearly the equilibrium  $(0, 0)$  is not stable if  $\varepsilon \neq 0$ . On the other hand, if  $U(x_1)$  is a smooth real valued analytic function, satisfying  $U(0) = 0$ ,  $U(x_1) > 0$  for  $x_1 \neq 0$ , then it is well known that for  $\varepsilon > 0$  the Newton equation

$$\dot{x} = \begin{pmatrix} 0 & \varepsilon \\ 0 & 0 \end{pmatrix} x - \begin{pmatrix} 0 \\ U'(x_1) \end{pmatrix}$$

is stable at  $(0, 0)$ .

Another example, which enters in the framework considered here, is given by a conservative (i.e. Hamiltonian) dynamical system perturbed by a friction term.

Next, we illustrate the influence of the real nilpotent perturbations in the realm of the normal form theory and in the general theory of singular partial differential equations. Consider the linear PDE

$$(x_1 + \varepsilon x_2) \partial_{x_1} u + x_2 \partial_{x_2} u - \sqrt{2} x_3 \partial_{x_3} u - u = f(x),$$

where  $\varepsilon \in \mathbb{R}$ , and  $f$  stands for a convergent power series in a neighborhood of the origin of  $\mathbb{R}^3$ , at least quadratic at  $x = 0$ . Such equations appear in the so-called homological equations for the reduction to Poincaré–Dulac linear normal form of systems of analytic ODEs having an equilibrium at the origin. It turns out that in the semisimple case

$\varepsilon = 0$  we can solve the equation in the space of convergent power series while for  $\varepsilon \neq 0$  (i.e., when a nontrivial Jordan block appears) the equation is solvable only formally, namely, divergent solutions appear. One is led in a natural way to study the Gevrey index of divergent solutions.

## Complex and Real Jordan Canonical Forms

We start by revisiting the notion of complex and real canonical Jordan forms.

Recall that each linear map is decomposed uniquely into the sum of a semisimple map and a nilpotent one. We state a classical result in linear algebra

**Lemma 1** Let  $A$  be an  $n \times n$  matrix with real or complex entries. Then it is uniquely decomposed as

$$A = A_s + A_{\text{nil}}, \quad (3)$$

where  $A_s$  is semisimple (i.e., diagonalizable over  $\mathbb{C}$ ) and  $A_{\text{nil}}$  is nilpotent, i.e.,  $A_{\text{nil}}^r = 0_{n \times n}$  for some positive integer  $r$ . Here  $0_{n \times n}$  stands for the zero  $n \times n$  matrix.

Denote by  $\text{spec}(A) = \{\lambda_1, \dots, \lambda_m\} \subset \mathbb{C}$  the set of all eigenvalues of  $A$  counted with their multiplicity.

The complex Jordan canonical form (JCF) is defined by the following assertion cf. Gantmacher (1959):

**Theorem 2** Let  $A$  be an  $n \times n$  matrix over  $\mathbb{K}$ ,  $\mathbb{K} = \mathbb{C}$  or  $\mathbb{K} = \mathbb{R}$ . Then there exist positive integers  $m, k_1, \dots, k_m$ ,  $m \geq n$ ,  $k_1 + \dots + k_m = n$  and a matrix  $S \in GL(n; \mathbb{C})$  such that

$$S^{-1}AS = J_A := \begin{pmatrix} \lambda_1^A I_{k_1} + N_{k_1} & 0_{k_1 \times k_2} & \dots & 0_{k_1 \times k_p} \\ 0_{k_2 \times k_1} & \lambda_2^A I_{k_2} + N_{k_2} & \dots & 0_{k_2 \times k_3} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{k_m \times k_1} & 0_{k_m \times k_2} & \dots & \lambda_m^A I_{k_m} + N_{k_m} \end{pmatrix} \quad (4)$$

where  $\lambda_1^A, \dots, \lambda_m^A$  are the eigenvalues of  $A$ , which need not all be distinct, and  $N_r$ , when  $r \geq 2$ , stands for the square  $r \times r$  matrix

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (5)$$

with the convention  $N_1 = 0$ . Moreover, if  $A_{\text{nil}} \neq 0$ , i.e.,  $k_j \geq 2$  for at least one  $j \in \{1, \dots, m\}$ , then for every  $\varepsilon \in \mathbb{C} \setminus 0$  the matrix  $S(\varepsilon) \in GL(n; \mathbb{C}) = \text{diag}\{S_1(\varepsilon), \dots, S_m(\varepsilon)\}$ ,  $S_j(\varepsilon) = 1$  if  $k_j = 1$ ,  $S_j(\varepsilon) = \text{diag}\{1, \dots, \varepsilon^{k_j-1}\}$ , provided  $k_j \geq 2$ ,  $j = 1, \dots, m$ , satisfies the identity

$$S^{-1}(\varepsilon)AS(\varepsilon) = J_A(\varepsilon) = \begin{pmatrix} \lambda_1^A I_{k_1} + \varepsilon N_{k_1} & 0_{k_1 \times k_2} & \dots & 0_{k_1 \times k_p} \\ 0_{k_2 \times k_1} & \lambda_2^A I_{k_2} + \varepsilon N_{k_2} & \dots & 0_{k_2 \times k_3} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{k_m \times k_1} & 0_{k_m \times k_2} & \dots & \lambda_m^A I_{k_m} + \varepsilon N_{k_m} \end{pmatrix}. \quad (6)$$

One may define in an obvious way another JCF (lower triangular) replacing  $J$  by its transposed  $J^T$ .

Note that  $\lambda_k^A$ ,  $k = 1, \dots, m$ , are not necessarily distinct.

If  $\mu$  is an eigenvalue of  $A$  with algebraic multiplicity  $d$ , i.e., it is a zero of multiplicity  $d$  of the characteristic polynomial

$$P_A(\lambda) = |A - \lambda I|,$$

one can have different Jordan block structures. For example, a  $3 \times 3$  matrix with a triple eigenvalue  $\mu$  can be reduced to one of the three JCF:

- the matrix  $\mu I_3$ , i.e.,  $\mu$  does not admit nontrivial Jordan blocks;
- $\begin{pmatrix} \mu & 1 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{pmatrix}$ ;
- $\begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{pmatrix}$

For higher dimensions the description of all JCF becomes more involved, cf. (Arnold 1983; Gantmacher 1959).

**Remark 3** We can choose  $|\varepsilon|$  arbitrarily small but never zero if the nilpotent part is nonzero. The columns of the conjugating matrix  $S$  are formed by eigenvectors and generalized eigenvectors. The smallness of  $|\varepsilon|$  leads in a natural way to view the presence of nontrivial nilpotent parts as a perturbation.

Next, if  $A$  is a real matrix, using the real and the imaginary parts of the eigenvectors for the complex eigenvalues, one introduces the real JCF.

**Theorem 4** Let  $A \in M_{n \times n}(\mathbb{R})$ . Then there exist nonnegative integers  $p, q$ ,  $1 \leq p + 2q \leq n$ ,  $p$  (if  $p \geq 1$ ) positive integers  $k_1, \dots, k_p$ ,  $q$  (if  $q \geq 1$ ) positive integers  $\ell_1, \dots, \ell_q$  satisfying  $k_1 + \dots + k_p + 2(\ell_1 + \dots + \ell_q) = n$ , and  $S \in SL(n; \mathbb{R})$ , such that

$$S^{-1}AS = \begin{pmatrix} J_A^R & 0_{k \times 2\ell} \\ 0_{2\ell \times k} & J_A^C \end{pmatrix} \quad (7)$$

with

$$J_A^R = \begin{pmatrix} \lambda_1^A I_{k_1} + \varepsilon_1^A N_{k_1} & 0_{k_1 \times k_2} & \dots & 0_{k_1 \times k_p} \\ 0_{k_2 \times k_1} & \lambda_2^A I_{k_2} + \varepsilon_2^A N_{k_2} & \dots & 0_{k_2 \times k_3} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{k_p \times k_1} & 0_{k_p \times k_2} & \dots & \lambda_p^A I_{k_p} + \varepsilon_p^A N_{k_p} \end{pmatrix}, \quad (8)$$

for some  $\lambda_j \in \mathbb{R}, j = 1, \dots, p$ , and

$$J_A^C = \begin{pmatrix} D_1^A & 0_{2\ell_1 \times 2\ell_2} & \dots & 0_{2\ell_1 \times 2\ell_q} \\ 0_{2\ell_2 \times 2\ell_1} & D_2^A & \dots & 0_{2\ell_2 \times 2\ell_q} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{2\ell_q \times 2\ell_1} & 0_{2\ell_q \times 2\ell_2} & \dots & D_q^A \end{pmatrix}, \quad (9)$$

with  $D_\mu^A$  being  $2\ell_\mu \times 2\ell_\mu$  matrices, written as  $\ell_\mu \times \ell_\mu$  block matrices of  $2 \times 2$  matrices of the following  $2 \times 2$  block matrix form:

$$D_\mu^A = \begin{pmatrix} \alpha_\mu & -\beta_\mu \\ \beta_\mu & \alpha_\mu \end{pmatrix} I_{\ell_\mu}(2) + \begin{pmatrix} \gamma_\mu^A & -\delta_\mu^A \\ \delta_\mu^A & \gamma_\mu^A \end{pmatrix} N_\mu(2), \quad (10)$$

for some  $\alpha_\mu, \beta_\mu \in \mathbb{R}, \beta_\mu \neq 0$ , with  $\alpha_k \pm \beta_k i \in \text{spec}(A)$ . Here,  $I_k(2) = \text{diag}\{I_2, \dots, I_2\}$  denotes the  $2k \times 2k$  matrix written as a  $k \times k$  matrix with  $2 \times 2$  block matrices while  $N_k(2)$  stands for the following  $2k \times 2k$  nilpotent matrix written as  $k \times k$  matrix with  $2 \times 2$  block matrices as entries:

$$N_k(2) = \begin{pmatrix} 0_{2 \times 2} & I_2 & 0_{2 \times 2} & \dots & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} & I_2 & \dots & 0_{2 \times 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & \dots & I_2 \\ 0_{2 \times 2} & 0_{2 \times 2} & 0_{2 \times 2} & \dots & 0_{2 \times 2} \end{pmatrix} \quad (11)$$

and  $0_{r \times s}$  stands for the zero  $r \times s$  matrix.

The smallness of the parameter  $\varepsilon$  and the explicit form of the conjugating matrices  $S(\varepsilon)$  are instrumental in showing some useful estimates for the study of the dynamics of the linear maps  $A$  which are not semisimple. Let  $r(A) := \max\{|\lambda| : \lambda \in \text{spec}(A)\}$  (the spectral radius of  $A$ ). Then the following estimate, useful in different branches of Dynamical Systems, holds (cf. (Hasselblatt and Katok 2003))

**Lemma 5** For every  $\eta > 0$  there exists a norm in  $\mathbb{R}^n$  such that  $\|A\| \leq r(A) + \eta$ . Moreover, if

$\{\lambda \in \text{spec}(A) : |\lambda| = r(A) \text{ do not admit Jordan blocks}\}$

one has  $\|A\| = r(A)$  for some norm. In particular, if  $A$  is semisimple, the last conclusion holds.

Next, consider the linear autonomous systems of ordinary differential equations

$$\dot{x} = Ax. \quad (12)$$

where  $A \in M_n(\mathbb{R})$ . We recall that if  $x(0) = \xi \in \mathbb{R}^n$ , then the unique solution is defined by

$$x(t) = \exp(tA)\xi := \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} \xi, \quad (13)$$

e.g., cf. Arnold (1983), Coddington and Levinson (1955).

We exhibit an assertion where the structure of the real nilpotent perturbation  $A_{\text{nil}}$  in the linear part plays a crucial role for the stability for  $t \geq 0$  of the solutions of the linear system (12). We recall that the origin is stable if for every  $\varepsilon > 0$  one can find  $\delta > 0$  such that  $\|\xi\| < \delta$  implies  $\|\exp(tA)\xi\| < \varepsilon$  for  $t \geq 0$ .

**Proposition 6** The zero solution of (12) is stable for  $t \geq 0$  if and only if the following two conditions hold:

- (i)  $\text{spec}(A) \subset \{\lambda \in \mathbb{C}; \text{Re } \lambda \leq 0\}$ ;
- (ii) if  $\lambda \in \text{spec}(A)$  and  $\text{Re } \lambda = 0$  then  $\lambda$  does not admit a nontrivial Jordan block.

On the other hand, the origin is asymptotically stable for  $t \rightarrow +\infty$  if and only if

$$\text{spec}(A) \subset \{\lambda \in \mathbb{C}; \text{Re } \lambda < 0\}.$$

The proof is straightforward in view of the explicit formula for the exponent of the matrix  $\exp(tA)$  by means of the Jordan canonical form.

In particular, we get

**Corollary 7** Let  $A$  be a real matrix such that all eigenvalues lie on the pure imaginary axis. Then the zero solution of (12) is stable if and only if  $A$  is semisimple (i.e.,  $A_{\text{nil}} = 0$ ).

## Nilpotent Perturbation and Formal Normal Forms of Vector Fields and Maps Near a Fixed Point

Normal form theory (originating back to Poincaré's thesis) has proven to be one of the most useful tools for the local analysis of dynamical systems near an equilibrium (singular) point  $x^0$  for autonomous systems of ODE

$$\dot{x} = X(x) \quad (14)$$

or the associated vector field

$$\tilde{X}(x) = \langle X(x), \partial_x \rangle = \sum_{j=1}^n X_j(x) \partial_{x_j} \quad (15)$$

(see (Arnold 1983; Belitskii 1979; Bruno 1971; Cicogna and Gaeta 1999; Gaeta and Walcher 2005; Gaeta and Walcher 2006; Sanders et al. 2007), and the references therein). Without loss of generality (after a translation) one may assume that  $x^0$  coincides with the origin and write

$$\begin{aligned} X(x) &= Ax + R(x), \quad A = \nabla X(0), \\ R(x) &= O(|x|^2), \quad |x| \rightarrow 0. \end{aligned} \quad (16)$$

Denote by  $\text{spec}(A) = \{\lambda_1, \dots, \lambda_n\}$  the spectrum of  $A$ . The basic idea, going back to Poincaré, is to find a (formal) change of the coordinates defined as a (at least) quadratic perturbation of the identity

$$x = u(y) = y + v(y), \quad (17)$$

which transforms  $\tilde{X}$  into a new vector field  $\tilde{Y}(y)$  which has a “simpler” form (Poincaré–Dulac normal form).

The original idea of Poincaré regards the possibility to linearize  $\tilde{X}$ , i.e.,  $\tilde{Y}(y) = Ay$ . Straightforward calculations show that the linearization of  $\tilde{X}$  means that  $v(y)$  satisfies (at least formally) a system of first order semilinear partial differential equations, called the system of the homological (difference) equations

$$L_A v(y) = R(y + v(y)) \quad (18)$$

In fact, it is the system above where the first substantial technical difficulty appears if the nilpotent part  $A_{\text{nil}}$  is not zero, i.e., the matrix  $A$  is not diagonalizable. Indeed, if  $A$  is diagonalizable and we choose (after a linear change of the variables in  $\mathbb{C}^n$ )  $A = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ , then the system (18) is written as

$$\sum_{k=1}^n \lambda_k \partial_{y_k} v_j(y) - \lambda_j v_j = R_j(y + v(y)), \quad (19)$$

$$j = 1, \dots, n$$

We recall that  $\tilde{X}$  (or  $\text{spec}(A)$ ) is said to be in the Poincaré domain (respectively, Siegel domain) if the convex hull of  $\{\lambda_1, \dots, \lambda_n\}$  in the complex plane does not contain (respectively, contains) 0. Further,  $\text{spec}(A)$  is called nonresonant iff

$$\langle \lambda, \alpha \rangle - \lambda_j \neq 0, \quad (20)$$

$$j = 1, \dots, n, \quad \alpha \in \mathbb{Z}_+^n(2),$$

where  $\langle \lambda, \alpha \rangle = \sum_{j=1}^n \lambda_j \alpha_j$ ,  $\mathbb{Z}_+^n(2) = \{\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n : |\alpha| := \alpha_1 + \dots + \alpha_n \geq 2\}$ .

By the Poincaré–Dulac theorem, if  $\text{spec}(A)$  is in the Poincaré domain, then there are at most finitely many resonances and there exists a convergent transformation (in some neighborhood of the singular point) which reduces  $\tilde{X}$  to a (finitely resonant) normal form.

**Theorem 8** Let the linear part  $A$  of the complex (respectively, real) field above be nonresonant. Then the vector field is formally linearizable by a complex (respectively, real) transformation.

Denote by

$$\text{Res}_j^A = \{\alpha \in \mathbb{Z}_+^n(2) : \lambda_j = \langle \lambda, \alpha \rangle\},$$

$$j = 1, \dots, n.$$

Clearly the nonresonance hypothesis is equivalent to  $\text{Res}_j^A = \emptyset$  for  $j = 1, \dots, n$ .

Additional technical complications appear if we consider real vector fields.

**Theorem 9** Every formal vector field with a singular point at the origin is transformed by a formal complex change of variables to a field of the form

$$\langle J_A z, \partial_z \rangle + \sum_{j=1}^n \sum_{\alpha \in \text{Res}_j^A} q_{j;\alpha} z^\alpha \partial_{z_j} \quad (21)$$

The coefficients  $q_{j;\alpha}$  may be complex even though the original vector field is real.

We note that if the linear part is nilpotent, i.e.,  $\text{spec}(A) = \{0\}$ , then the theorem above gives no simplification. In that case Belitskii (1979) has classified completely the formal normal forms.

Let  $A$  be a nilpotent matrix (i.e. the semisimple part  $A_s$  is the zero matrix). The Poincaré–Dulac NF does not provide any information. The following theorem is due to Belitskii (1979) (see also Arnold and Ilyashenko (1988)).

**Theorem 10** Let  $A$  be a nilpotent matrix and let  $X(x)$  be a formal vector field with a linear part given by  $Ax$ . Then  $X$  is transformed by a formal complex change of variables  $x \mapsto z$  to a field of the form

$$\langle J_A z, \partial_z \rangle + \langle B(z), \partial_z \rangle \quad (22)$$

with  $B(z)$  being at least quadratic near the origin, where the nonlinear vector field  $\langle B(z), \partial_z \rangle$  commutes with  $\langle J_A^* z, \partial_z \rangle$ . (Here  $^*$  stands for the Hermitian conjugation).

We point out that another important problem is the computation of the normal form. Here the presence of the Jordan blocks leads to substantial difficulties.

The description and the computation of nilpotent normal forms, based on an algebraic approach and the systematic use of the theory of invariance, with particular emphasis on the equivalence between different normal forms, has been developed in a body of papers (see (Cushman and Sanders 1990; Murdock 2002; Murdock and Sanders 2007), and the references therein).

Finally, we mention that nilpotent perturbations appear in the classification and Casimir invariants of Lie–Poisson brackets that are formed by Lie algebra extensions for physical systems admitting Hamiltonian structure to such brackets (e.g. cf. Thiffeault and Morison (2000)).

## Loss of Gevrey Regularity in Siegel Domains in the Presence of Jordan Blocks

The convergence question in the Siegel domain is more difficult since small divisors appear. In a fundamental paper Bruno (1971) succeeded in

proving a deep result of the following type: a formal normal form is convergent under an (optimal) arithmetic condition on the small divisors  $|\langle \lambda, \alpha \rangle - \lambda_j|^{-1}$  and a condition on the formal normal form, called the  $A$  condition. It should be pointed out that while in the original paper (Bruno 1971) the condition  $A$  allows in some cases non-trivial Jordan blocks of the linear part, in the subsequent works the linear part  $A$  is required to be semisimple (diagonalizable).

We recall that for the finitely smooth and  $C^\infty$  local normal forms the presence of the real nilpotent part does not influence the convergence: e.g., in the Sternberg theorem (cf. (Sternberg 1958)) the small divisors and the presence of Jordan blocks play no role (cf. (Belitskii 1979), where many other references can be found).

Little is known about convergence–divergence problems in the analytic category if  $\text{spec}(A)$  is in the Siegel domain and  $A$  is *not* semisimple.

Even in the cases where the presence of the nilpotent part does not influence the assertions, the proofs become more involved. Here we outline various aspects of normal forms when nilpotent perturbations are present. We stress especially the real case.

The combined influence of the Jordan blocks and the small divisors on the convergence of the formal linearizing transformation for analytic vectorfields in  $\mathbb{R}^3$  has been studied in Gramchev (2002) (see also (Yoshino 1999) for some examples). Let  $n = 3$  and consider a non semisimple linear part  $A \in GL(3; \mathbb{K})$  satisfying  $\text{spec}(A) = \{\lambda, \mu, \mu\}$ ,  $\lambda \neq \mu$ . This means that we can reduce  $A$  to the  $\varepsilon$  Jordan normal form

$$A_\varepsilon = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & \varepsilon \\ 0 & 0 & \mu \end{pmatrix}, \quad \varepsilon \neq 0. \quad (23)$$

We recall that one can make  $|\varepsilon|$  arbitrarily small by linear change of the variables (but never 0). Then  $\text{spec}(A)$  is nonresonant and in the Siegel domain iff

$$\rho := \frac{\lambda}{\mu} < 0, \quad \rho \notin \mathbb{Q}. \quad (24)$$

The typical example is a real vector field with a linear part given by

$$A_\varepsilon^0 = \begin{pmatrix} \rho & 0 & 0 \\ 0 & 1 & \varepsilon \\ 0 & 0 & 1 \end{pmatrix}, \quad \varepsilon \neq 0. \quad (25)$$

One observes that such real vector fields are nonresonant and hyperbolic and therefore, by the Chen theorem, linearizable by smooth transformations.

We recall that an irrational number  $\rho$  is said to be diophantine of order  $\tau > 0$ , and write  $\rho \in D(\tau)$ , if there exist  $C > 0$  such that

$$\min_{p \in \mathbb{Z}} |q\rho + p| \geq \frac{C}{q^\tau}, \quad q \in \mathbb{N}. \quad (26)$$

By a classical result in number theory  $D(\tau) \neq \emptyset$  iff  $\tau \geq 1$ . An irrational number  $\rho$  is called Liouville iff it is not diophantine.

Given an irrational number  $\rho$  we set

$$\begin{aligned} \tau_0 &= \tau_0(\rho) \\ &= \inf\{\tau > 0 : \text{such that } \rho \in D(\tau)\}, \text{ before} \end{aligned} \quad (27)$$

with the convention  $\tau_0 = +\infty$  if  $\rho$  is a Liouville number.

**Theorem 11** Let  $\text{spec}(A)$  be in the Siegel domain. Then  $L_A$  is not solvable in the space of convergent power series, namely, we can find RHS  $f$  which is analytic but the unique formal power series solution is divergent. Moreover, we can always find a convergent RHS  $f$  such that the unique formal solution  $u$  satisfies

$$u \notin \bigcup_{1 \leq \sigma < 2 + \tau_0} G_f^\sigma(\mathbb{K}^3) \quad (28)$$

In particular, if  $\tau_0 = +\infty$ , then

$$u \notin G^\sigma, \quad \sigma \geq 1. \quad (29)$$

In case a diophantine condition is satisfied, estimates on the Gevrey character of the divergent power series are derived.

The nonsolvability of the LHE in the spaces of the convergent power series does not exclude a priori that the vector field is linearizable by



analytic transformations. However, using a fundamental result of Pérez Marco (2001) one can state the following assertion for hyperbolic vector fields which are not in the Poincaré domain.

**Theorem 12** Let  $A$  be a real  $n \times n$  matrix which is hyperbolic, not semisimple and is not in the Poincaré domain. Let  $R(x) = (R_1(x), \dots, R_n(x))$  be a real valued analytic function near the origin, at least quadratic near  $x = 0$ , i.e.  $R(0) = 0$ ,  $\nabla R(0) = 0$ . Then there exists  $\eta_0 > 0$  such that for almost all  $\eta \in ]0, \eta_0]$  in the sense of the Lebesgue measure, the vector field defined by

$$X_\varepsilon(x) = Ax + \eta R(x) \quad (30)$$

is not linearizable by convergent transformations.

**Remark 13** In fact, the result can be made more precise, using the notion of capacity instead of the Lebesgue measure and allowing polynomial dependence on  $\eta$  (cf. (Pérez Marco 2001)).

Another direction where the presence of real nilpotent perturbations in the linear parts presents challenging obstacles is the study of the dynamics in a neighborhood of a fixed point carried out via a normalization up to finite order and the issue of optimal truncation

$$\sum_{|\alpha| \leq N_{\text{opt}}} u_\alpha x^\alpha$$

of normal form transformations for analytic vector fields. In an impressive work Iooss and Lombardi (2005) demonstrate in particular that for large classes of real analytic vector fields with semisimple linear parts one can truncate the formal normal form transformation for  $|x| \leq \delta$ ,  $\delta_0 > 0$  arbitrarily small, in such a way that the reminder in the normal form  $R(x)$  satisfies the following estimates

$$\sup_{|x| \leq \delta} |R_{N_{\text{opt}}}(x)| \leq M \delta^2 \exp\left(-\frac{w}{\delta^b}\right) \quad (31)$$

where  $b = 1 + \tau$ . Here either  $\tau > n - 1$ , in which case  $\tau \neq 0$  is the diophantine index of the small

divisor type estimates modulo the resonance set, namely for some  $\gamma > 0$  the following estimates hold for the eigenvalues  $\lambda_1, \dots, \lambda_n$  of  $A$

$$|\langle \lambda, \alpha \rangle - \lambda_j| \geq \frac{\gamma}{|\alpha|^\tau}, \quad \text{if } \langle \lambda, \alpha \rangle - \lambda_j \neq 0 \quad (32)$$

for  $j = 1, \dots, n$ ,  $\alpha \in \mathbb{Z}_+^n(2)$ , or  $\tau = 0$  and  $\lambda$  satisfies the nonresonant type estimates

$$|\langle \lambda, \alpha \rangle - \lambda_j| \geq \gamma, \quad \text{if } \langle \lambda, \alpha \rangle - \lambda_j \neq 0 \quad (33)$$

for  $j = 1, \dots, n$ ,  $\alpha \in \mathbb{Z}_+^n(2)$ .

The question of the validity of such results for analytic vector fields with non semisimple linearization is far more intricate. Iooss and Lombardi (2005) give two examples of non-semisimple linearizations (nilpotent perturbations of size 2 and 3) for which the result is still true. The question remains totally open for other non-semisimple linearizations.

## First-Order Singular Partial Differential Equations

This section deals with the study of formal power series solutions to singular linear first-order partial differential equations with analytic coefficients of the form

$$\sum_{j=1}^d a_j(x) \partial_{x_j} u(x) + b(x) u(x) = f(x), \quad (34)$$

where  $a_j(x)$  (with  $j = 1, \dots, d$ ),  $b(x)$  and the right-hand side  $f(x)$  are analytic in a neighborhood of the origin of  $\mathbb{C}^d$ , and  $a_j(0) = 0$  for  $j = 1, \dots, d$ .

In an interesting paper Hibino (1999) allows a Jacobian matrix  $\nabla a(0)$  without a Poincaré condition. More precisely, the Jacobi matrix at the origin can be reduced via a conjugation with a nonsingular matrix  $S$  to the following form: for some nonnegative integers,  $m, d, p$ , and  $d$  positive integers  $k_j \geq 2$  (with  $j = 1, \dots, d$ ), if  $d \geq 1$ , such that  $m + r_1 + \dots + r_d + p = n$ , one can write as follows:



$$S^{-1}\nabla a(0)S = \begin{pmatrix} A & & & \\ & N_{r_1} & & \\ & & \ddots & \\ & & & N_{r_d} \\ & & & & 0_{p \times p} \end{pmatrix} \quad (35)$$

where  $N_r$ ,  $r \geq 2$ , stands for the  $r \times r$  nilpotent Jordan block (5) and  $A$  is an  $m \times m$  satisfying the Poincaré condition (the convex hull in  $\mathbb{C}$  of the eigenvalues  $\lambda_1, \dots, \lambda_m$  of  $A$  does not contain the origin, provided  $m \geq 1$ ).

One observes that the hypothesis  $d \geq 1$  implies that  $\nabla a(0)$  does not satisfy the Poincaré condition.

The fundamental hypothesis on the zero order term  $b(x)$  reads as follows

$$\left| \sum_{r=1}^m \lambda \alpha_r + b(0) \right| \neq 0, \quad \alpha \in \mathbb{Z}_+^m \quad \text{if} \quad m \geq 1 \quad (36)$$

$$|b(0)| \neq 0 \quad \text{if} \quad m = 0 \quad (37)$$

In particular, by (37) one gets that necessarily  $b(0) \neq 0$ .

It should be stressed that the classes of singular partial differential equations above do not capture the systems of homological equations when small divisors occur, but they outline some interesting features in the presence of nontrivial Jordan blocks even in the lack of small divisors phenomena.

Set  $\tau_0 = \max_{\ell=1, \dots, d} \tau_\ell$  if  $d \geq 1$ . Then the main result in (Hibino 1999) reads as follows

**Theorem 14** Under the conditions (35), (36), and (37) for every RHS

$$f(x) = \sum_{\alpha \in \mathbb{Z}_+^n} f_\alpha x^\alpha$$

which converges in a neighborhood of the origin the eq. (34) has a unique formal solution which belongs to the formal Gevrey space  $G_f^\sigma(\mathbb{K}^n)$  with

$$\sigma = \begin{cases} 2\tau_0 & \text{if } d \geq 1 \\ 2 & \text{if } d = 0, p \geq 1 \\ 1 & \text{if } d = p = 0 \end{cases} \quad (38)$$

The Gevrey index is determined by a Newton polyhedron, a generalization of the notion of the Newton polygon for singular ordinary differential equations. The arguments of the proof rely on subtle Gevrey combinatorial estimates.

Extensions for first-order quasilinear singular partial differential equations are done by Hibino (2006).

It should be pointed out that the loss of Gevrey regularity comes not only from the nilpotent Jordan blocks, but from the nonlinear (at least quadratic) terms in  $a(x)$  as well. Indeed, for the equation

$$(1 - x_1^2 \partial_{x_1} - x_2^2 \partial_{x_2})u(x) = x_1 + x_2$$

the unique formal solution is defined as follows

$$u(x) = \sum_{j=0}^{\infty} (j-1)! (x_1^j + x_2^j)$$

cf. (Hibino 1999). For further investigations on the loss of Gevrey regularity for solutions of singular ordinary differential equations of irregular type see Gramchev and Yoshino (2007). For characterizations of the Borel summability of a divergent formal power series solution of classes of first-order linear singular partial differential equation of nilpotent type see (Hibino 2003) and the references therein.

Solvability in classical Sobolev spaces and Gevrey spaces for linear systems of singular partial differential equations with real coefficients in  $\mathbb{R}^n$  with nontrivial real Jordan blocks are derived by Gramchev and Tolis (2006).

### Normal Forms for Real Commuting Vector Fields with Linear Parts Admitting Nontrivial Jordan Blocks

The main goal of this section is to exhibit the influence of nontrivial linear nilpotent parts for the simultaneous reduction to convergent normal forms of commuting vector fields with a common fixed point.

Consider a family of commuting  $n \times n$  matrices  $A_1, \dots, A_d$ . If all matrices are semi-simple, then they can be simultaneously diagonalized over  $\mathbb{C}$  or put into a block-diagonal form over  $\mathbb{R}$ , if  $A_1, \dots, A_d$  are real, by a linear transformation  $S$  (e.g., (Cicogna and Gaeta 1999; Gantmacher 1959)). This property plays a crucial role in the study of the normal forms of commuting vector fields with semi-simple linear parts (cf. (Stolovitch 2000, 2005)).

However, if the matrices have nontrivial nilpotent parts, then it is not possible to transform simultaneously  $A_1, \dots, A_d$  in Jordan canonical forms. This is a consequence of the characterization of the centralizer of a matrix in a JCF (see (Arnold 1983; Gantmacher 1959)).

Apparently the first examples of simultaneous reduction to (formal) normal forms of commuting vector fields with non semisimple linear parts are due to Cicogna and Gaeta (1999) for two commuting vector fields using the set up of the symmetries. More precisely, first, the definition of Semisimple Joint Normal Form (SJNF) is introduced: let

$$X = \langle f(x), \partial_x \rangle, \quad Y = \langle g(x), \partial_x \rangle,$$

$A = \nabla f(0)$ ,  $B = \nabla g(0)$  and  $f(x) = Ax + F(x)$ ,  $g(x) = Bx + G(x)$ . Then  $X$  and  $Y$  are said to be in SJNF if both  $F$  and  $G$  belong to  $\text{Ker}(\mathcal{A}_s) \cap \text{Ker}(\mathcal{B}_s)$ . Here  $\mathcal{A}_s$  stands for the homological operator associated to the semisimple part  $A_s$  of  $A$ .

Next,  $X, Y$  are in  $X$ -Joint NF iff

$$F \in \text{Ker}(\mathcal{A}^+) \cap \text{Ker}(\mathcal{B}_s),$$

and  $g \in \text{Ker}(\mathcal{A}_s) \cap \text{Ker}(\mathcal{B}_s)$ .

Clearly in  $Y$ -JNF the role of  $f$  and  $g$  is reversed. The main assertion is:

**Theorem 15** Let  $[X, Y] = 0$ . Then  $X$  and  $Y$  can be reduced to a SJNF by means of a formal change of the variables. They can also be reduced (formally) to  $X$ -Joint or  $Y$ -Joint NF.

### Example 16

$$A = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & \varepsilon \\ 0 & 0 & \lambda \end{pmatrix} \quad B = \begin{pmatrix} p & 0 & q \\ r & s & t \\ 0 & 0 & s \end{pmatrix}, \quad (39)$$

where  $\lambda, \varepsilon, p, q, r, s, t \in \mathbb{C} \setminus 0$ . Then  $A$  and  $B$  commute but it is impossible, in general, to reduce  $B$  to the Jordan block structure of  $A$ , preserving that of  $A$ .

We point out that the problem of finding conditions guaranteeing simultaneous reduction of commuting matrices to Jordan normal forms is related to questions in the Lie group theory (e.g., see Chap. IV of (Cicogna and Gaeta 1999) and the references therein).

Next, we outline recent results on simultaneous reductions to normal forms of commuting analytic vector fields admitting nontrivial Jordan blocks in their linear parts following Yoshino and Gramchev (2008).

Let  $\mathbb{K}$  be  $\mathbb{K} = \mathbb{C}$  or  $\mathbb{K} = \mathbb{R}$ , and  $B = \infty$ ,  $B = \omega$  or  $B = k$  for some  $k > 0$ . Let  $\mathcal{G}_B^n$  denote a  $d$ -dimensional Lie algebra of germs at  $0 \in \mathbb{K}^n$  of  $C^B$  vector fields vanishing at 0. Let  $\rho$  be a germ of singular infinitesimal  $\mathbb{K}^d$ -actions of class  $C^B$  ( $d \geq 2$ )

$$\rho : \mathbb{K}^d \rightarrow \mathcal{G}_B^n. \quad (40)$$

Denote by  $\text{Act}^B(\mathbb{K}^d : \mathbb{K}^n)$  the set of germs of singular infinitesimal  $\mathbb{K}^d$ -actions of class  $C^B$  at  $0 \in \mathbb{K}^n$ . By choosing a basis  $e_1, \dots, e_d \in \mathbb{K}^n$ , the infinitesimal action can be identified with a  $d$ -tuple of germs at 0 of commuting vector fields  $X^j = \rho(e_j)$ ,  $j = 1, \dots, d$  (cf. (Dumortier and Roussarie 1980; Stolovitch 2000, 2005; Zung 2002)). We can define, in view of the commutativity relation, the action

$$\begin{aligned} \tilde{\rho} : \mathbb{K}^d \times \mathbb{K}^n &\rightarrow \mathbb{K}^n, \\ \tilde{\rho}(s; z) &= X_{s_1}^1 \circ \dots \circ X_{s_d}^d(z) = X_{s_{\sigma_1}}^{\sigma_1} \circ \dots \circ X_{s_{\sigma_d}}^{\sigma_d}(z), \quad (41) \\ s &= (s_1, \dots, s_d), \end{aligned}$$

for all permutations  $\sigma = (\sigma_1, \dots, \sigma_d)$  of  $\{1, \dots, d\}$ , where  $X_t^j$  denotes the flow of  $X^j$ . We

denote by  $\rho_{\text{lin}}$  the linear action formed by the linear parts of the vector fields defining  $\rho$ .

A natural question is to investigate necessary and sufficient conditions for the linearization of  $\rho$  (allowing nilpotent perturbations in the linear parts) namely, whether there exists a  $C^B$  diffeomorphism  $g$  preserving 0 such that  $g$  conjugates  $\tilde{\rho}$  and  $\tilde{\rho}_{\text{lin}}$

$$\tilde{\rho}(s; g(z)) = g(\tilde{\rho}_{\text{lin}}(s, z)) \quad (s, z) \in \mathbb{K}^d \times \mathbb{K}^n. \quad (42)$$

It is well known (e.g., cf. (Katok and Katok 1995)) that there exists a positive integer  $m \leq n$  such that  $\mathbb{K}^n$  is decomposed into a direct sum of  $m$  linear subspaces invariant under all  $A^\ell = \nabla X_\ell(0)$  ( $\ell = 1, \dots, d$ ):

$$\begin{aligned} \mathbb{K}^n &= \mathbb{I}^{s_1} + \dots + \mathbb{I}^{s_m}, \dim \mathbb{I}^{s_j} = s_j, \\ j &= 1, \dots, m, s_1 + \dots + s_m = n. \end{aligned} \quad (43)$$

The matrices  $A^1, \dots, A^d$  can be simultaneously brought into upper triangular form, and we write again  $A^\ell$  for the matrices

$$A^\ell = \begin{pmatrix} A_1^\ell & 0_{s_1 \times s_2} & \dots & 0_{s_1 \times s_m} \\ 0_{s_2 \times s_1} & A_2^\ell & \dots & 0_{s_2 \times s_m} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{s_m \times s_1} & 0_{s_m \times s_2} & \dots & A_m^\ell \end{pmatrix}, \quad (44)$$

$\ell = 1, \dots, d.$

If  $\mathbb{K} = \mathbb{C}$ , the matrix  $A_j^\ell$  is given by

$$A_j^\ell = \begin{pmatrix} \lambda_j^\ell & A_{j,12}^\ell & \dots & A_{j,1s_j}^\ell \\ 0 & \lambda_j^\ell & \dots & A_{j,2s_j}^\ell \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_j^\ell \end{pmatrix}, \quad (45)$$

with  $\lambda_j^\ell, A_{j,v\mu}^\ell \in \mathbb{C}$ ,  $\ell = 1, \dots, d$ ,  $j = 1, \dots, m$ . On the other hand, if  $\mathbb{K} = \mathbb{R}$ , then we have, for every  $1 \leq j \leq m$  two possibilities: firstly, all  $A_j^\ell$  ( $\ell = 1, \dots, d$ ) are given by (45) with  $\lambda_j^\ell \in \mathbb{R}$ . Secondly,  $s_j = 2\tilde{s}_j$  is even and  $A_j^\ell$  is a  $\tilde{s}_j \times \tilde{s}_j$  square block matrix given by

$$A_j^\ell = \begin{pmatrix} R_2(\lambda_j^\ell, \mu_j^\ell) & A_{\ell j}^{12} & \dots & A_{\ell j}^{1s_j} \\ 0 & R_2(\lambda_j^\ell, \mu_j^\ell) & \dots & A_{\ell j}^{2s_j} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_2(\lambda_j^\ell, \mu_j^\ell) \end{pmatrix}, \quad (46)$$

$\ell = 1, \dots, d,$

where

$$R_2(\lambda, \mu) := \begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}, \quad \lambda, \mu \in \mathbb{R}, \quad (47)$$

and  $A_{\ell j}^{rs}$  are appropriate real matrices.

Following the decomposition (45) (respectively, (46)) we define  $\tilde{\lambda}^j$  by

$$\tilde{\lambda}^k = {}^t(\lambda_1^k, \dots, \lambda_m^k) \in \mathbb{K}^m, \quad k = 1, \dots, d. \quad (48)$$

Then we assume

$$\tilde{\lambda}^1, \dots, \tilde{\lambda}^d \text{ are linearly independent in } \mathbb{K}^m. \quad (49)$$

One can easily see that (49) is invariantly defined.

By (44) we define

$$\vec{\lambda}_j = {}^t(\lambda_j^1, \dots, \lambda_j^d) \in \mathbb{K}^d, \quad j = 1, \dots, m, \quad (50)$$

and

$$A_m := \{\vec{\lambda}_1, \dots, \vec{\lambda}_m\}. \quad (51)$$

We define the cone  $\Gamma[A_m]$  by

$$\Gamma[A_m] = \left\{ \sum_{j=1}^m t_j \vec{\lambda}_j \in \mathbb{K}^d; t_j \geq 0, j = 1, \dots, m, \sum_{j=1}^m t_j \neq 0 \right\}. \quad (52)$$

**Definition 17** A  $\mathbb{K}^d$ -action  $\rho$  is called a Poincaré morphism if there exists a basis  $A_m \subset \mathbb{K}^m$  such that  $\Gamma[A_m]$  is a proper cone in  $\mathbb{K}^m$ , namely it does not contain a straight real line. If the condition is

not satisfied, then, we say that the  $\mathbb{K}^d$ -action is in a Siegel domain.

Note that the definition is invariant under the choice of the basis  $\mathcal{A}_m$ .

**Remark 18** The geometric definition above is equivalent to the notion of Poincaré morphism given by Stolovitch (Definition 6.2.1 in (Stolovitch 2000)).

Next, we need to introduce the notion of simultaneous resonance. For  $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{K}^m$ ,  $\beta = (\beta_1, \dots, \beta_m) \in \mathbb{K}^m$ , we set  $\langle \alpha, \beta \rangle = \sum_{v=1}^m \alpha_v \beta_v$ . For a positive integer  $k$  we define  $\mathbb{Z}_+^m(k) = \{\alpha \in \mathbb{Z}_+^m; |\alpha| \geq k\}$ . Put

$$\omega_j(\alpha) = \sum_{v=1}^d \left| \left\langle \tilde{\lambda}^v, \alpha \right\rangle - \lambda_j^v \right|, \quad (53)$$

$$j = 1, \dots, m,$$

$$\omega(\alpha) = \min \{\omega_1(\alpha), \dots, \omega_m(\alpha)\}. \quad (54)$$

**Definition 19** The cone  $\mathcal{A}_m$  is called simultaneously nonresonant (or, in short  $\rho$  is simultaneously nonresonant), if

$$\omega(\alpha) \neq 0, \quad \forall \alpha \in \mathbb{Z}_+^m(2). \quad (55)$$

If (55) does not hold, then  $\mathcal{A}_m$  is said to be simultaneously resonant.

Clearly, the simultaneously nonresonant condition (55) is invariant under a change of the basis  $\mathcal{A}_m$ .

The next assertion provides a geometrically invariant condition guaranteeing that the simultaneous reduction to normal form does not depend on (small) nilpotent perturbation of the linear part.

**Theorem 20** Let  $\rho$  be a Poincaré morphism. Then  $\rho$  is conjugated to a polynomial action by a convergent change of variables.

**Remark 21** As a corollary of Theorem 20 for vector fields having linear parts with nontrivial Jordan blocks one obtains generalizations of results for the existence of convergent normal forms for analytic vector fields admitting

symmetries cf. (Bambusi et al. 1998; Cicogna and Gaeta 1999; Cicogna and Walcher 2002).

**Example 22** Let  $\rho$  be a  $\mathbb{R}^2$ -action in  $\mathbb{R}^n$ ,  $n \geq 4$  with  $m = 3$ . Choose a basis  $\mathcal{A}_2$  of  $\mathbb{R}^3$  such that

$$\mathcal{A}_2 = \{ {}^t(1, 1, v), {}^t(0, 1, \mu) \}, \quad v, \mu \in \mathbb{R}. \quad (56)$$

By (52),  $\Gamma[\mathcal{A}_2]$  is generated by the set of vectors  $\{(1, 0), (1, 1), (v, \mu)\}$ . Hence the action is a Poincaré morphism if and only if these vectors generate a proper cone, namely  $(v, \mu)$  is not in the set  $\{(v, \mu) \in \mathbb{R}^2; v \leq \mu \leq 0\}$ . We note that the interesting case is  $\mu < v \leq 0$ , where every generator in (56) is in a Siegel domain.

Next, given a two-dimensional Lie algebra, choose a basis  $X_1, X_2$  with linear parts  $A_j \in GL(4; \mathbb{C})$  satisfying  $\text{spec}(A_1) = \{1, 1, v, v\}$  and  $\text{spec}(A_2) = \{0, 1, \mu, \mu\}$ , respectively, where  $v < \mu < 0$ ,  $(v, \mu) \notin \mathbb{Q}^2$ , and

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & v & 0 \\ 0 & 0 & 0 & v \end{pmatrix}, \quad (57)$$

$$A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \mu & \varepsilon \\ 0 & 0 & 0 & \mu \end{pmatrix},$$

where  $\varepsilon \neq 0$ .

We show a refinement of the divergence result in Gevrey classes in (Yoshino and Gramchev 2008) for the solution  $v$  of the overdetermined systems of linear homological equations  $L_j v := \nabla v(x) A_j x - A_j v = f^j$  ( $j = 1, 2$ ), with the compatibility conditions for the RHS (see (Yoshino and Gramchev 2008) for more details).

**Theorem 23** Let  $1/2 \leq \tau_0 < \infty$ . Then there exists

$$E_0 \subset \left\{ (v, \mu) \in (\mathbb{R} \setminus \mathbb{Q})^2; v < \mu < 0, \right. \\ \left. v \text{ does not satisfy the Bruno condition} \right\}$$

with the density of continuum such that for every  $(v, \mu) \in E_0$ , there exists analytic  $f = {}^t(f^1, f^2)$

$\in (\mathbb{C}_2^4\{x\})^2$ , satisfying the compatibility condition for the overdetermined system and such that the unique formal solution  $v(x)$  is not in  $\cup_{1 \leq \sigma < 2+\tau} G^\sigma(\mathbb{C}^4)$ . Moreover, for every analytic  $f$  we can find  $C > 0$  such that the unique formal solution satisfies the anisotropic Gevrey estimates

$$|v_\alpha| \leq C^{|\alpha|+1} (\alpha_3 + \alpha_4)^{(1+\tau)\alpha_4}, \alpha \in \mathbb{Z}_+^4(2), \quad (58)$$

### Analytic Maps near a Fixed Point in the Presence of Jordan Blocks

As for the real analytic local diffeomorphisms preserving the origin in  $\mathbb{R}^n$ , one has in fact to deal necessarily with hyperbolic maps (cf. (Arnold and Ilyashenko 1988)).

Recall first the complex analytic case cf. (Arnold 1983; Arnold and Ilyashenko 1988). Let  $\Phi(x)$  be a biholomorphic map of  $\mathbb{C}^n$  preserving the origin and  $\Phi'(0)$  the Jacobian matrix at the origin. Denote by  $\text{spec}(\Phi'(0)) = \{\lambda_1, \dots, \lambda_n\}$  the spectrum of  $\Phi'(0)$ . Clearly  $\lambda_j \neq 0$  for all  $j = 1, \dots, n$ . We define the set of all resonance multi indices of  $\Phi$  (actually it depends only on  $\text{spec}(\Phi'(0))$ ) as follows:

$$\begin{aligned} \text{Res}[\lambda_1, \dots, \lambda_n] &= \bigcup_{j=1}^n \text{Res}^j[\lambda_1, \dots, \lambda_n] \\ \text{Res}^j[\lambda_1, \dots, \lambda_n] &= \{\alpha \in \mathbb{Z}_+^n(2) : \lambda^\alpha - \lambda_j = 0\}; \end{aligned} \quad (59)$$

where  $\mathbb{Z}_+^n(k) := \{\alpha \in \mathbb{Z}_+^n : |\alpha| \geq k\}$ ,  $\lambda = (\lambda_1, \dots, \lambda_n)$  and  $\lambda^\alpha = \lambda_1^{\alpha_1} \dots \lambda_n^{\alpha_n}$  for  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n$ .

Given  $A \in GL(n, \mathbb{C})$  define  $O[A]$  as the germ of all local complex analytic diffeomorphisms (biholomorphic maps)  $\Phi(x)$  of  $\mathbb{C}^n$  with 0 as a fixed point and such that  $\Phi'(0)$  is in the  $GL(n, \mathbb{C})$ -conjugacy class of  $A$ . We stress that if  $A$  is diagonalizable (semisimple) then the germ is determined by the spectrum of  $A$ , i.e., by  $n$  nonzero complex numbers  $\lambda_1, \dots, \lambda_n$ .

We will say that  $\Phi(x)$  is formally linearizable if there exists a formal series  $u(x) = x + \sum_{\alpha \in \mathbb{Z}_+^n(2)} u_\alpha x^\alpha$  such that

$$u^{-1} \circ \Phi \circ u = \Phi'(0) \quad \text{formally in } (\mathbb{C}[x])^n, \quad (60)$$

where  $\mathbb{C}[x]$  stands for the set of all formal power series with complex coefficients. It is well known (the Poincaré–Dulac theorem, cf. (Arnold 1983)) that under the nonresonance hypothesis  $\text{Res}[\lambda_1, \dots, \lambda_n] = \emptyset$ , i.e.,

$$\lambda^\alpha - \lambda_j \neq 0, \quad \alpha \in \mathbb{Z}_+^n(2), \quad j = 1, \dots, n, \quad (61)$$

$\Phi$  is formally linearizable. In fact, (61) is a necessary and sufficient condition in order that every holomorphic  $\Phi(x)$  with  $\text{spec}\Phi'(0) = \{\lambda_1, \dots, \lambda_n\}$  is formally linearizable. We refer to Gramchev and Walcher (2005) for formal and algebraic aspects of normal forms of maps.

The formal solution of (60) involves expressions of the form  $(\lambda^\alpha - \lambda_j)^{-1}$ ; so when  $\inf_\alpha |\lambda^\alpha - \lambda_j| = 0$  for some  $j \in \{1, \dots, n\}$ , the convergence of  $u$  becomes a subtle question.

One of the main problems, starting from the pioneering work of Siegel (1942), has been (and still is) to find general conditions which guarantee that  $u(x)$  converges, i.e., that  $\Phi(x)$  is linearizable. We recall the state of the art of this subject. If the linear part  $\Phi'(0)$  is semisimple (i.e.,  $\Phi'(0)$  has no nontrivial Jordan blocks), then it is well known that for convergence we need arithmetic (Diophantine) conditions on

$$\omega(m) := \min_{2 \leq |\alpha| \leq m, 1 \leq j \leq n} |\lambda^\alpha - \lambda_j|, \quad m \in \mathbb{Z}_+(2). \quad (62)$$

We refer for the history and references to the survey paper by Herman (1987). The best condition that implies linearizability for all maps with a given semisimple linear part is due to Bruno (1971), and can be expressed as (following Herman, p. 143 in (1987))

$$\sum_{k=1}^{\infty} 2^{-k} \ln(\omega^{-1}(2^{k+1})) < \infty. \quad (63)$$

If one assumes the Poincaré condition

$$\max_{1 \leq j \leq n} |\lambda_j| < 1 \quad \text{or} \quad \max_{1 \leq j \leq n} |\lambda_j| > 1, \quad (64)$$

then  $\Phi$  is always analytically equivalent to its linear part  $\Phi'(0)$ , provided the nonresonance hypothesis holds. More generally, (64) implies that there are finitely many resonances and, according to the Poincaré–Dulac theorem, we can find a local biholomorphic change of the variables  $u$  bringing  $\Phi$  to normal form

$$(u^{-1} \circ \Phi \circ u)(x) = \Phi'(0)x + P_{\text{res}}(x), \quad (65)$$

where the remainder  $P_{\text{res}}(x)$  is a polynomial map containing only resonant terms.

Little is known about the (non)linearizability of  $\Phi$  in the analytic category if the Poincaré condition doesn't hold and the matrix  $\Phi'(0)$  has at least one nontrivial Jordan block. When  $n = 2$  and  $A$  has a double eigenvalue  $\lambda_1 = \lambda_2$ ,  $|\lambda_1| = 1$  and a nontrivial Jordan block, then in general  $\Phi$  is not linearizable. This result is contained in Proposition 3, p. 143 in (Herman 1987), which is a consequence of results of Ilyashenko (1979) and Yoccoz (the latter proved in 1978; for published proofs we refer to Appendix, pp. 86–87 in (Yoccoz 1995)). It is not difficult to extend this negative result to  $\mathbb{C}^n$ ,  $n \geq 3$ .

In a recent paper of DeLatte and Gramchev (2002) biholomorphic maps in  $\mathbb{C}^n$ ,  $n = 3$  and  $n = 4$  having a single nontrivial Jordan block of the linear part have been studied. We mention also the paper of Abate (2000), where nondiagonalizable discrete holomorphic dynamical systems have been investigated using geometrical tools.

One observes that in the real case, as the matrix  $\Phi'(0)$  is real, the nonresonance condition excludes eigenvalues on the unit circle, i.e., the map is hyperbolic.

As a corollary from results of Delatte and Gramchev (2002) on nonsolvability of the LHE in the presence of Jordan blocks in the linear parts of biholomorphic maps preserving the origin of  $\mathbb{C}^3$  and the fundamental results of Pérez Marco (2001) one gets readily the following assertion for hyperbolic analytic maps preserving the origin in  $\mathbb{R}^n$  and having nondiagonalizable linear parts at the origin.

**Theorem 24** Let  $A$  be a real  $n \times n$  matrix such that its eigenvalues are nonresonant and lie outside the unit circle in  $\mathbb{C}$ , and  $A$  is neither expansive

nor contractive (i.e., the Poincaré condition (64) is not satisfied). Let  $R(x) = (R_1(x), \dots, R_n(x))$  be a real valued analytic function near the origin, at least quadratic near  $x = 0$ , i.e.  $R(0) = 0$ ,  $\nabla R(0) = 0$ . Then there exists  $\eta_0 > 0$  such that for almost all  $\eta \in ]0, \eta_0]$  in the sense of the Lebesgue measure, the local analytic diffeomorphism defined by

$$\Phi(x) = Ax + \eta R(x) \quad (66)$$

is not linearizable by convergent transformations.

## Weakly Hyperbolic Systems and Nilpotent Perturbations

The presence of nondiagonalizable matrices appear as a challenging problem in the framework of the well-posedness of the Cauchy problem for evolution partial differential equations. In order to illustrate the main features we focus on linear hyperbolic systems with constant coefficients in space dimension one

$$\partial_t u + A \partial_x u + Bu = 0, \quad t > 0, \quad x \in \mathbb{R} \quad (67)$$

$$u(0, x) = u^0(x) \quad (68)$$

where  $A$  and  $B$  are real  $m \times m$  matrices,  $u = (u_1, \dots, u_m)$  stands for a vector-valued smooth function. One is interested in the  $C^\infty$  well-posedness of the Cauchy problem (67), (68), namely, for every initial data  $u^0 \in (C^\infty(\mathbb{R}))^m$  there exists a unique solution  $u \in C^\infty(\mathbb{R}^2)$  of (67), (68).

We recall that the system is hyperbolic if the characteristic equation

$$|A - \lambda I| = 0 \quad (69)$$

has  $m$  real solutions  $\lambda_1, \dots, \lambda_m$ . If the roots are distinct, the system is called strictly hyperbolic. The strict hyperbolicity implies that  $A$  is diagonalizable, which in turn implies that, after a linear change of variables, one can assume  $A$  to be diagonal and reduce essentially the problem to  $m$  first-order scalar equations. In that case the Cauchy problem is well-posed in the classical sense, namely for every smooth initial data  $u^0$  (it is

enough  $u^0 \in C^1(\mathbb{R})$ ) there exists a unique smooth solution  $u(t, x)$  to the Cauchy problem. In fact, it is enough to require that  $A$  is semisimple (i.e., allowing multiple eigenvalues but excluding nilpotent parts) in order to have well-posedness (e.g., cf. the book of Taylor (1981) and the references therein for more details on strictly hyperbolic systems with variable coefficients, and more general set-up in the framework of pseudo differential operators).

If the system is hyperbolic but not strictly hyperbolic (it is called also weakly hyperbolic), it means that the matrix  $A$  has multiple eigenvalues. Here the influence of the Jordan block structure is decisive and one has non existence theorems in the  $C^\infty$  category unless one imposes additional restrictions on the lower-order term  $B$ . In many applications one encounters weakly hyperbolic systems. One example is in the so-called water waves problem, concerning the motion of the free surface of a body of an incompressible irrotational fluid under the influence of gravity (see (Craig 1987) and the references therein), where for the linearized system one encounters a  $2 \times 2$  matrix of the type

$$A = \begin{pmatrix} c & \varkappa \\ 0 & c \end{pmatrix},$$

where  $c$  stands for the velocity, and  $\varkappa$  is a nonzero real number.

The assertions, even in the seemingly simple model cases, are not easy to state in simple terms. The first results on such systems are due to Kajitani (1979). The classification problem was completely settled for the so-called hyperbolic systems of constant multiplicities by Vaillant (1999) using subtle linear algebra arguments with multiparameter dependence if the space dimension is greater than 1.

We illustrate the assertions for  $m = 2$  and  $m = 3$ , where the influence of the nilpotent part is somewhat easier to describe in details. In what follows we rewrite the results in (Kajitani 1979; Vaillant 1999) by means of the Jordan block structures.

The case  $m = 2$  is easy.

**Proposition 25** Let

$$A = \begin{pmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{pmatrix}, \quad (70)$$

for some  $\lambda_0 \in \mathbb{R}$ . Then the Cauchy problem is  $C^\infty$  well posed if  $b_{21} = 0$ , with

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Let the  $3 \times 3$  real matrix  $A$  have a triple eigenvalue  $\lambda_0$  and be not semisimple. Then we are reduced (modulo conjugation with an invertible matrix) to two possibilities: either the JCF of  $A$  is a maximal Jordan block

$$\begin{pmatrix} \lambda_0 & 1 & 0 \\ 0 & \lambda_0 & 1 \\ 0 & 0 & \lambda_0 \end{pmatrix}, \quad (71)$$

or the degenerate case of one  $2 \times 2$  and one  $1 \times 1$  elementary Jordan blocks

$$A = \begin{pmatrix} \lambda_0 & 1 & 0 \\ 0 & \lambda_0 & 0 \\ 0 & 0 & \lambda_0 \end{pmatrix}, \quad (72)$$

**Proposition 26** Let  $n = 3$ . Then the following assertions hold:

- (i) let  $A$  be defined by (71). Then the Cauchy problem is well-posed in  $C^\infty$  if the entries of the matrix

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

satisfy the identities

$$b_{31} = b_{21} + b_{32} = b_{11} - b_{13} = 0; \quad (73)$$

- (ii) suppose that  $A$  is given by (72). The well-posedness in  $C^\infty$  holds iff



$$b_{21} = b_{21}b_{32} = 0. \quad (74)$$

In an interesting work, Petkov (1979), using real Jordan block structures depending on parameters and reduction to normal forms of matrices depending on parameters (cf. (Arnold 1971)), derived canonical micro local forms for the full symbol of a pseudo differential system with real characteristics of constant multiplicity and applies them to study the propagation of singularities of solutions of certain systems.

More generally, the Cauchy problem for hyperbolic systems with multiple characteristics have been studied by various authors where (implicitly) conditions on the nilpotent perturbations and the lower order term are imposed (e.g., cf. (Bove and Nishitani 2003; Kajitani 1979; Yamahara 2000), and the references therein).

It would be interesting to write down such conditions in terms of conditions on the nilpotent perturbations on the principal part and the lower-order terms.

Finally, we mention also the work of Ghedamsi, Gourdin, Mechab, and Takeuchi (2002), concerned with the Cauchy problem for Schrödinger-type systems with characteristic roots of multiplicity two admitting nontrivial Jordan blocks.

## Bibliography

- Abate M (2000) Diagonalization of nondiagonalizable discrete holomorphic dynamical systems. *Am J Math* 122: 757–781
- Arnold VI (1971) Matrices depending on parameters. *Uspekhi Mat Nauk* 26:101–114 (in Russian); *Russ Math Surv* 26:29–43
- Arnold VI (1983) Geometrical methods in the theory of ordinary differential equations. Springer, New York
- Arnold VI, Ilyashenko YU (1988) In: Anosov DV, Arnold VI (eds) *Encyclopedia of Math Sci*, vol 1. Dynamical systems I. Springer, New York, pp 1–155
- Bambusi D, Cicogna G, Gaeta G, Marmo G (1998) Normal forms, symmetry and linearization of dynamical systems. *J Phys A* 31:5065–5082
- Belitskii GR (1978) Equivalence and normal forms of germs of smooth mappings. *Uspekhi Mat Nauk* 33: 95–155, 263 (in Russian); *Russ Math Surv* 33:107–177
- Belitskii GR (1979) Normal forms, invariants, and local mappings. *Naukova Dumka*, Kiev (in Russian)
- Bove A, Nishitani T (2003) Necessary conditions for hyperbolic systems. II. *Jpn J Math (NS)* 29:357–388
- Bruno AD (1971) The analytic form of differential equations. *Tr Mosk Mat O-va* 25:119–262; (1972) 26: 199–239 (in Russian); See also (1971) *Trans Mosc Math Soc* 25:131–288; (1972) 26:199–239
- Bruno AD, Walcher S (1994) Symmetries and convergence of normalizing transformations. *J Math Anal Appl* 183: 571–576
- Chen KT (1965) Diffeomorphisms:  $C^\infty$ -realizations of formal properties. *Am J Math* 87:140–157
- Cicogna G, Gaeta G (1999) Symmetry and perturbation theory in nonlinear dynamics. Lecture notes in physics. New series m: monographs, vol 57. Springer, Berlin
- Cicogna G, Walcher S (2002) Convergence of normal form transformations: the role of symmetries. (English summary) *Symmetry and perturbation theory. Acta Appl Math* 70:95–111
- Coddington EA, Levinson N (1955) *Theory of ordinary differential equations*. McGraw-Hill, New York
- Craig W (1987) Nonstrictly hyperbolic nonlinear systems. *Math Ann* 277:213–232
- Cushman R, Sanders JA (1990) A survey of invariant theory applied to normal forms of vector fields with nilpotent linear part. In: Stanton D (ed) *Invariant theory and tableaux*. IMA vol Math Appl, vol 19. Springer, New York, pp 82–106
- DeLatte D, Gramchev T (2002) Biholomorphic maps with linear parts having Jordan blocks: linearization and resonance type phenomena. *Math Phys Electron J* 8(2): 1–27
- Dumortier F, Roussarie R (1980) Smooth linearization of germs of  $R^2$ -actions and holomorphic vector fields. *Ann Inst Fourier Grenoble* 30:31–64
- Gaeta G, Walcher S (2005) Dimension increase and splitting for Poincaré-Dulac normal forms. *J Nonlinear Math Phys* 12(1):327–342
- Gaeta G, Walcher S (2006) Embedding and splitting ordinary differential equations in normal form. *J Differ Equ* 224:98–119
- Gantmacher FR (1959) *The theory of matrices*, vols 1, 2. Chelsea, New York
- Ghedamsi M, Gourdin D, Mechab M, Takeuchi J (2002) Équations et systèmes du type de Schrödinger à racines caractéristiques de multiplicité deux. *Bull Soc R Sci Liège* 71:169–187
- Gramchev T (2002) On the linearization of holomorphic vector fields in the Siegel Domain with linear parts having nontrivial Jordan blocks. In: Abenda S, Gaeta G, Walcher S (eds) *Symmetry and perturbation theory*, Cala Gonone, 16–22 May 2002. World Sci. Publ, River Edge, pp 106–115
- Gramchev T, Tolis E (2006) Solvability of systems of singular partial differential equations in function spaces. *Integral Transform Spec Funct* 17: 231–237
- Gramchev T, Walcher S (2005) Normal forms of maps: formal and algebraic aspects. *Acta Appl Math* 85: 123–146
- Gramchev T, Yoshino M (2007) Normal forms for commuting vector fields near a common fixed point. In: Gaeta G, Vitolo R, Walcher S (eds) *Symmetry and*

- perturbation theory, Oltranto, 2–9 June 2007. World Sci Publ, River Edge, pp 203–217
- Hasselblatt B, Katok A (2003) A first course in dynamics: with a panorama of recent developments. Cambridge University Press, Cambridge
- Herman M (1987) Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of  $C^n$  near a fixed point. VIIIth international congress on mathematical physics, Marseille 1986. World Sci Publ, Singapore, pp 138–184
- Hibino M (1999) Divergence property of formal solutions for first order linear partial differential equations. Publ Res Inst Math Sci 35:893–919
- Hibino M (2003) Borel summability of divergent solutions for singular first order linear partial differential equations with polynomial coefficients. J Math Sci Univ Tokyo 10:279–309
- Hibino M (2006) Formal Gevrey theory for singular first order quasi-linear partial differential equations. Publ Res Inst Math Sci 42:933–985
- Il'yashenko Y (1979) Divergence of series reducing an analytic differential equation to linear form at a singular point. Funct Anal Appl 13:227–229
- Iooss G, Lombardi E (2005) Polynomial normal forms with exponentially small remainder for vector fields. J Differ Equ 212:1–61
- Kajitani K (1979) Cauchy problem for non-strictly hyperbolic systems. Publ Res Inst Math 15:519–550
- Katok A, Katok S (1995) Higher cohomology for Abelian groups of toral automorphisms. Ergod Theory Dyn Syst 15:569–592
- Murdock J (2002) On the structure of nilpotent normal form modules. J Differ Equ 180:198–237
- Murdock J, Sanders JA (2007) A new transvectant algorithm for nilpotent normal forms. J Differ Equ 238: 234–256
- Pérez Marco R (2001) Total convergence or small divergence in small divisors. Commun Math Phys 223: 451–464
- Petkov VM (1979) Microlocal forms for hyperbolic systems. Math Nachr 93:117–131
- Ramis J-P (1984) Théorèmes d'indices Gevrey pour les équations différentielles ordinaires. Mem Am Math Soc 48:296
- Rodino L (1993) Linear partial differential operators in Gevrey spaces. World Science, Singapore
- Sanders JA (2005) Normal form in filtered Lie algebra representations. Acta Appl Math 87:165–189
- Sanders JA, Verhulst F, Murdock J (2007) Averaging methods in nonlinear dynamical systems, 2nd edn. Applied mathematical sciences, vol 59. Springer, New York
- Siegel CL (1942) Iteration of analytic functions. Ann Math 43:607–614
- Sternberg S (1958) The structure of local homeomorphisms. II, III. Am J Math 80:623–632, 81:578–604
- Stolovitch L (2000) Singular complete integrability. Publ Math IHES 91:134–210
- Stolovitch L (2005) Normalisation holomorphe d'algèbres de type Cartan de champs de vecteurs holomorphes singuliers. Ann Math 161:589–612
- Taylor M (1981) Pseudodifferential operators. Princeton mathematical series, vol 34. Princeton University Press, Princeton
- Thiffeault J-L, Morison PJ (2000) Classification and Casimir invariants of Lie–Poisson brackets. Physica D 136:205–244
- Vaillant J (1999) Invariants des systèmes d'opérateurs différentiels et sommes formelles asymptotiques. Jpn J Math (NS) 25:1–153
- Yamahara H (2000) Cauchy problem for hyperbolic systems in Gevrey class. A note on Gevrey indices. Ann Fac Sci Toulouse Math 19:147–160
- Yoccoz J-C (1995) A remark on Siegel's theorem for nondiagonalizable linear part. Manuscript, 1978; See also Théorème de Siegel, nombres de Bruno e polynômes quadratic. Astérisque 231:3–88
- Yoshino M (1999) Simultaneous normal forms of commuting maps and vectorfields. In: Degasperis A, Gaeta G (eds) Symmetry and perturbation theory SPT 98, Rome, 16–22 December 1998. World Scientific, Singapore, pp 287–294
- Yoshino M, Gramchev T (2008) Simultaneous reduction to normal forms of commuting singular vector fields with linear parts having Jordan blocks. Ann Inst Fourier (Grenoble) 58:263–297
- Zung NT (2002) Convergence versus integrability in Poincaré–Dulac normal form. Math Res Lett 9:217–228

## Perturbation Theory for PDEs

Dario Bambusi

Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

### Article Outline

Glossary

Definition of the Subject

Introduction

The Hamiltonian Formalism for PDEs

Normal Form for Finite Dimensional Hamiltonian Systems

Normal Form for Hamiltonian PDEs: General Comments

Normal Form for Resonant Hamiltonian PDEs and Its Consequences

Normal Form for Nonresonant Hamiltonian PDEs

Non Hamiltonian PDEs

Extensions and Related Results

Future Directions

Bibliography

### Glossary

**Perturbation theory** The study of a dynamical systems which is a perturbation of a system whose dynamics is known. Typically the unperturbed system is linear or integrable.

**Normal form** The normal form method consists of constructing a coordinate transformation which changes the equations of a dynamical system into new equations which are as simple as possible. In Hamiltonian systems the theory is particularly effective and typically leads to a very precise description of the dynamics.

**Hamiltonian PDE** A Hamiltonian PDE is a partial differential equation (abbreviated PDE) which is equivalent to the Hamilton equation of a suitable Hamiltonian function. Classical examples are the nonlinear wave equation, the

Nonlinear Schrödinger equation, and the Kortweg–de Vries equation.

**Resonance vs. Non-Resonance** A frequency vector  $\{\omega_k\}_{k=1}^n$  is said to be non-resonant if its components are independent over the relative integers. On the contrary, if there exists a non-vanishing  $K \in \mathbb{Z}^n$  such that  $\omega \cdot K = 0$  the frequency vector is said to be resonant. Such a property plays a fundamental role in normal form theory. Non-resonance typically implies stability.

**Actions** The action of a harmonic oscillator is its energy divided by its frequency. It is usually denoted by  $I$ . The typical issue of normal form theory is that in nonresonant systems the actions remain approximatively unchanged for very long times. In resonant systems there are linear combinations of the actions with such properties.

**Sobolev space** Space of functions which have weak derivatives enjoying suitable integrability properties. Here we will use the spaces  $H^s$ ,  $s \in \mathbb{N}$  of the functions which are square integrable together with their first  $s$  weak derivatives.

### Definition of the Subject

Perturbation theory for PDEs is a part of the qualitative theory of differential equations. One of the most effective methods of perturbation theory is the normal form theory which consists of using coordinate transformations in order to describe the qualitative features of a given or generic equation. Classical normal form theory for ordinary differential equations has been used all along the last century in many different domains, leading to important results in pure mathematics, celestial mechanics, plasma physics, biology, solid state physics, chemistry and many other fields.

The development of effective methods to understand the dynamics of partial differential equations is relevant in pure mathematics as well as in all the fields in which partial differential

equations play an important role. Fluid dynamics, oceanography, meteorology, quantum mechanics, and electromagnetic theory are just a few examples of potential applications. More precisely, the normal form theory allows one to understand whether a small nonlinearity can change the dynamics of a linear PDE or not. Moreover, it allows one to understand how the changes can be avoided or forced. Finally, when the changes are possible it allows to predict the behavior of the perturbed system.

## Introduction

The normal form method was developed by Poincaré and Birkhoff between the end of the nineteenth century and the beginning of the twentieth century. During the last 20 years the method has been successfully generalized to a suitable class of partial differential equations (PDEs) in finite volume (in the case of infinite volume dispersive effects appear and the theory is very different. See e.g. (Soffer and Weinstein 1999)). In this article we will give an introduction to this recent field. We will almost only deal with Hamiltonian PDEs, since on the one hand the theory for non Hamiltonian systems is a small variant of the one we will present here, and on the other hand most models are Hamiltonian.

We will start by a generalization of the Hamiltonian formalism to PDEs, followed by a review of the classical theory and by the actual generalization of normal form theory to PDEs.

In the next section we give a generalization of the Hamiltonian formalism to PDEs. The main new fact is that in PDEs the Hamiltonian is usually a smooth function, but the corresponding vector field is nonsmooth (it is an operator extracting derivatives). So the standard formalism has to be slightly modified (Bambusi and Giorgilli 1993; Chernoff and Marsden 1974; Kappeler and Pöschel 2003; Kuksin 1993; Marsden 1972; Weinstein 1969). Here we will present a version of the Hamiltonian formalism which is enough to cover the models of interest for local perturbation theory. To clearly illustrate the situation we will start the article with an introduction to the

Lagrangian and Hamiltonian formalism for the wave equation. This will lead to the introduction of the paradigm Hamiltonian which is usually studied in this context. This will be followed by a few results on the Hamiltonian formalism that are needed for perturbation theory.

Subsequently, we shortly present the standard Birkhoff normal form theory for finite dimensional systems. This is useful since all the formal aspects are equal in the classical case and the case of PDEs.

Then we come to the generalization of normal form theory to PDEs. In the present paper we will concentrate almost only on the case of 1-dimensional semilinear equations. This is due to the fact that the theory of higher dimensional and quasilinear equations is still quite unsatisfactory.

In PDEs one essentially meets two kinds of difficulties. The first one is related to the existence of non smooth vector fields. The second difficulty is due to the fact that in the infinite dimensional case there are small denominators which are much worse than in the finite dimensional one.

We first present the theory for completely resonant systems (Bambusi and Giorgilli 1993; Bambusi and Nekhoroshev 1998) in which the difficulties related to small denominators do not appear. It turns out that it is quite easy to obtain a normal form theorem for resonant PDEs, but the kind of normal form one gets is usually quite poor. In order to extract dynamical informations from the normal form one can only compute and study it explicitly. Usually this is very difficult. Nevertheless in some cases it is possible and leads to quite strong results. We will illustrate such a situation by studying a nonlinear Schrödinger equation (Bambusi 1999a; Bourgain 2000).

For the general case there is a theorem ensuring that a generic system admits at least one family of “periodic like trajectories” which are stable over exponentially long times (Bambusi and Nekhoroshev 2002). We will give its statement and an application to the nonlinear wave equation

$$u_{tt} - u_{xx} + \mu^2 u + f(u) = 0, \quad (1)$$

with  $\mu = 0$  and the Dirichlet boundary conditions on a segment (Paleari et al. 2001).

Then we turn to the case of nonresonant PDEs. The main difficulty is that small denominators accumulate to zero already at order 3. Such a problem has been overcome in (Bambusi 2003, 2008; Bambusi et al. 2007b; Bambusi and Grébert 2006; Grébert 2007) by taking advantage of the fact that the nonlinearities appearing in PDEs typically have a special form. In this case one can deduce a very precise description of the dynamics and also some interesting results of the kind of almost global existence of smooth solutions (Klainerman 1983). To illustrate the theory we will make reference to the nonlinear wave Eq. (1) with almost any  $\mu$ , and to the nonlinear Schrödinger equation.

Another aspect of the theory of close to integrable Hamiltonian PDEs concerns the extension of KAM theory to PDEs. We will not present it here. We just recall the most celebrated results which are those due to Kuksin (1987), Wayne (1990), Craig–Wayne (1993), Bourgain (1998, 2005a), Kuksin–Pöschel (1996), Eliasson–Kuksin (2006). All these results ensure the existence of families of quasiperiodic solutions, i.e. solutions lying on finite dimensional manifolds. We also mention the papers (Bourgain 1996a; Pöschel 2002) where some Cantor families of full dimension tori are constructed. We point out that in the dynamics on such  $\infty$ -dimensional tori the amplitude of oscillation of the linear modes decreases super exponentially with their index. A remarkable exception is provided by the paper (Bourgain 2005b) where the tori constructed are more “thick” (even if of course they lie on Cantor families).

On the contrary, the results of normal form theory describe solutions starting on opens subsets of the phase space, and do not have particularly strong localizations properties with respect to the index. The price one has to pay is that the description one gets turns out to be valid only over long but finite times.

Finally we point out a related research stream that has been carried on by Bourgain (Bourgain 1996a, b, 1997, 2000) who studied intensively the behavior of high Sobolev norms

in close to integrable Hamiltonian PDEs (see also (Bourgain and Kaloshin 2005)).

## The Hamiltonian Formalism for PDEs

### The Gradient of a Functional

**Definition 1** Consider a function  $f \in C^\infty(\mathcal{U}_s, \mathbb{R})$ ,  $\mathcal{U}_s \subset H^s(\mathbb{T})$  open,  $s \geq 0$  a fixed parameter and  $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$  is the 1-dimensional torus. We will denote by  $\nabla f(u)$  the gradient of  $f$  with respect to the  $L^2$  metric, namely the unique function such that

$$\langle \nabla f(u), h \rangle_{L^2} = df(u)h, \quad \forall h \in H^s \quad (2)$$

where

$$\langle u, v \rangle_{L^2} := \int_{-\pi}^{\pi} u(x)v(x)dx \quad (3)$$

is the  $L^2$  scalar product and  $df(u)$  is the differential of  $f$  at  $u$ . The gradient is a smooth map from  $H^s$  to  $H^{-s}$  (see e.g. (Bambusi 1999b)).

**Example 2** Consider the function

$$f(u) := \int_{-\pi}^{\pi} \frac{u_x^2}{2} dx, \quad (4)$$

which is differentiable as a function from  $H^s \rightarrow \mathbb{R}$  for any  $s \geq 1$ . One has

$$\begin{aligned} df(u)h &= \int_{-\pi}^{\pi} u_x h_x dx = \int_{-\pi}^{\pi} -u_{xx} h dx \\ &= \langle -u_{xx}, h \rangle_{L^2} \end{aligned} \quad (5)$$

and therefore in this case one has  $\nabla f(u) = -u_{xx}$ .

**Example 3** Let  $\mathcal{F} : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a smooth function and define

$$f(u) = \int_{-\pi}^{\pi} \mathcal{F}(u, u_x) dx \quad (6)$$

then the gradient of  $f$  coincides with the so called functional derivative of  $\mathcal{F}$ :

$$\nabla f \equiv \frac{\delta \mathcal{F}}{\delta u} := \frac{\delta \mathcal{F}}{\delta u} - \frac{\partial}{\partial x} \frac{\partial \mathcal{F}}{\partial u_x}. \quad (7)$$

### Lagrangian and Hamiltonian Formalism for the Wave Equation

Until Subsection “[Basic Elements of Hamiltonian Formalism for PDEs](#)” we will work at a formal level, without specifying the function spaces and the domains.

**Definition 4** Let  $L(u, \dot{u})$  be a Lagrangian function, then the corresponding Lagrange equations are given

$$\nabla_u L - \frac{d}{dt} \nabla_{\dot{u}} L = 0 \quad (8)$$

where  $\nabla_u L$  is the gradient with respect to  $u$  only, and similarly  $\nabla_{\dot{u}}$  is the gradient with respect to  $\dot{u}$ .

**Example 5** Consider the Lagrangian

$$L(u, \dot{u}) := \int_{-\pi}^{\pi} \left( \frac{\dot{u}^2}{2} - \frac{u_x^2}{2} - \mu^2 \frac{u^2}{2} - F(u) \right) dx. \quad (9)$$

then the corresponding Lagrange equations are given by (1) with  $f = -F$ .

Given a Lagrangian system with a Lagrangian function  $L$ , one defines the corresponding Hamiltonian system as follows.

**Definition 6** Consider the momentum  $v := \nabla_{\dot{u}} L$  conjugated to  $u$ ; assume that  $L$  is convex with respect to  $\dot{u}$ , then the Hamiltonian function associated to  $L$  is defined by

$$H(v, u) := [\langle v; \dot{u} \rangle_{L^2} - L(u, \dot{u})]_{\dot{u}=\dot{u}(u, v)}. \quad (10)$$

**Definition 7** Let  $H(v, u)$  be a Hamiltonian function, then the corresponding Hamilton equations are given by

$$\dot{v} = -\nabla_u H, \quad \dot{u} = \nabla_v H. \quad (11)$$

As in the finite dimensional case, one has that the Lagrange equations are equivalent to the Hamilton equation of  $H$ .

An elementary computation shows that for the wave equation one has  $v = \dot{u}$  and

$$H(v, u) = \int_{-\pi}^{\pi} \left( \frac{v^2 + u_x^2 + \mu^2 u^2}{2} + F(u) \right) dx \quad (12)$$

### Canonical Coordinates

Consider a Lagrangian system and let  $\mathbf{e}_k$  be an orthonormal basis of  $L^2$ , write  $u = \sum_k q_k \mathbf{e}_k$  and  $\dot{u} = \sum_k \dot{q}_k \mathbf{e}_k$ , then one has the following proposition.

**Proposition 8** The Lagrange Eqs. (8) are equivalent to

$$\frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} = 0 \quad (13)$$

**Proof** Taking the scalar product of (8) with  $\mathbf{e}_k$  one gets

$$\langle \mathbf{e}_k; \nabla_u L \rangle_{L^2} - \frac{d}{dt} \langle \mathbf{e}_k; \nabla_{\dot{u}} L \rangle_{L^2} = 0$$

but one has  $\langle \mathbf{e}_k; \nabla_u L \rangle = \frac{\partial L}{\partial q_k}$  and similarly for the other term. Thus the thesis follows.

This proposition shows that, once a basis has been introduced, the Lagrange equations have the same form as in the finite dimensional case.

In the Hamiltonian case exactly the same result holds. Precisely, denoting  $v := \sum_k p_k \mathbf{e}_k$  one has the following proposition.

**Proposition 9** The Hamilton equations of a Hamiltonian function  $H$  are equivalent to

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}, \quad \dot{q}_k = \frac{\partial H}{\partial p_k}. \quad (14)$$

In the case of the nonlinear wave equation, in order to get a convenient form of the equations, one can choose the Fourier basis. Such a basis is defined by

$$\widehat{e}_k := \begin{cases} \frac{1}{\sqrt{\pi}} \cos kx & k > 0 \\ \frac{1}{\sqrt{2\pi}} & k = 0 \\ \frac{1}{\sqrt{\pi}} \sin -kx & k < 0 \end{cases} \quad (15)$$

Thus the Hamiltonian (12) takes the form

$$H(p, q) = \sum_{k \in \mathbb{Z}} \frac{p_k^2 + \omega_k^2 q_k^2}{2} + \int_{-\pi}^{\pi} F\left(\sum_k q_k \widehat{e}_k(x)\right) dx, \quad (16)$$

where  $\omega_k^2 := k^2 + \mu^2$ . For the forthcoming developments it is worth to rescale the variables by defining

$$p'_k := \frac{p_k}{\sqrt{\omega_k}}, \quad q'_k := \sqrt{\omega_k} q_k, \quad (17)$$

so that, omitting primes, the Hamiltonian takes the form

$$H(p, q) = \sum_k \omega_k \frac{p_k^2 + q_k^2}{2} + H_P(p, q) \quad (18)$$

where  $H_P$  has a zero of order higher than 2. In the following we will always study systems of the form (18). Moreover, by relabeling the variables and the frequencies it is possible to reduce the problem to the case where  $k$  varies in  $\mathbb{N} \equiv \{1, 2, 3, \dots\}$ . This is what we will assume in developing the abstract theory.

**Example 10** An example of a different nature in which the Hamiltonian takes the form (18) is the nonlinear Schrödinger equation

$$-i\dot{\psi} = \psi_{xx} + f(|\psi|^2)\psi, \quad (19)$$

where  $f$  is a smooth function. Eq. (19) has the conserved energy functional

$$H(\psi, \overline{\psi}) := \int_{-\pi}^{\pi} \left( |\psi|^2 + F(|\psi|^2) \right) dx, \quad (20)$$

where  $F$  is such that  $F' = f$ . Introduce canonical coordinates  $(p_k, q_k)$  by

$$\psi = \sum_{k \in \mathbb{Z}} \frac{p_k + iq_k}{\sqrt{2}} \widehat{e}_k, \quad (21)$$

then the energy takes the form (18) with  $\omega_k = k^2$  and the NLS is equivalent to the corresponding Hamilton equations.

**Example 11** Consider the Kortweg–de Vries equation

$$u_t + u_{xxx} + uu_x = 0, \quad (22)$$

in the space of functions with zero mean value. The conserved energy is given by

$$H(u) = \int_{-\pi}^{\pi} \left( \frac{u_x^2}{2} + \frac{u^3}{6} \right) dx, \quad (23)$$

which again is also the Hamiltonian of the system. Canonical coordinates are here introduced by

$$u(x) = \sum_{k>0} \sqrt{k} (p_k \widehat{e}_k + q_k \widehat{e}_{-k}), \quad (24)$$

in which the Hamiltonian takes the form (18) with  $\omega_k = k^3$ .

**Remark 12** It is also interesting to study some of these equations with Dirichlet boundary conditions (DBC), typically on  $[0, \pi]$ . This will always be done by identifying the space of the functions fulfilling DBC with the space of the function fulfilling periodic boundary conditions on  $[-\pi, \pi]$  which are skew symmetric. Similarly, Neumann boundary conditions will be treated by identifying the corresponding functions with periodic even functions. In some cases (e.g. in Eq. (1) with DBC and an  $f$  which does not have particular symmetries) the equations do not extend naturally to the space of skew symmetric and this has some interesting consequences (see (Bambusi et al. 2007a, 2008)).

## Basic Elements of Hamiltonian Formalism for PDEs

A suitable topology in the phase space is given by a Sobolev like topology.



For any  $s \in \mathbb{R}$ , define the Hilbert space  $\ell_s^2$  of the sequences  $x \equiv \{x_k\}_{k \geq 1}$  with  $x_k \in \mathbb{R}$  such that

$$\|x\|_s^2 := \sum_k |k|^{2s} |x_k|^2 < \infty \quad (25)$$

and the phase spaces  $\mathcal{P}_s := \ell_s^2 \oplus \ell_s^2 \equiv z \ni (p, q) \equiv (\{p_k\}, \{q_k\})$ . In  $\mathcal{P}_s$  we will sometimes use the scalar product

$$\langle (p, q), (p^1, q^1) \rangle_s := \langle p, p^1 \rangle_{\ell_s^2} + \langle q, q^1 \rangle_{\ell_s^2}. \quad (26)$$

In the following we will always assume that

$$|\omega_k| \leq C|k|^d \quad (27)$$

for some  $d$ .

**Remark 13** Defining the operator  $A_0 : D(A_0) \rightarrow \mathcal{P}_s$  by  $A_0(p, q) = (\omega_k p_k, \omega_k q_k)$  one can write  $H_0 = \frac{1}{2} \langle A_0 z; z \rangle_0$ ,  $D(A_0) \supset \mathcal{P}_{s+d}$ .

Given a smooth Hamiltonian function  $\chi : \mathcal{P}_s \supset \mathcal{U}_s \rightarrow \mathbb{R}$ ,  $\mathcal{U}_s$  being an open neighborhood of the origin, we define the corresponding Hamiltonian vector field  $X_\chi : \mathcal{U}_s \mapsto \mathcal{P}_{-s}$  by

$$X_\chi \equiv \left( -\frac{\partial \chi}{\partial q_k}, \frac{\partial \chi}{\partial p_k} \right). \quad (28)$$

**Remark 14** Corresponding to a function  $\chi$  as above we will denote by  $\nabla \chi$  its gradient with respect to the  $\ell^2 \equiv \ell_0^2$  metric. Defining the operator  $J$  by  $J(p, q) := (-q, p)$  one has  $X_\chi = J \nabla \chi$ .

**Definition 15** The Poisson Bracket of two smooth functions  $\chi_1, \chi_2$  is formally defined by

$$\{\chi_1; \chi_2\} := d\chi_1 X_{\chi_2} \equiv \langle \nabla \chi_1; J \nabla \chi_2 \rangle_0. \quad (29)$$

**Remark 16** As the example  $\chi_1 = \sum_k k q_k$ ,  $\chi_2 := \sum_k k p_k$  shows, there are cases where the Poisson Bracket of two functions is not defined.

For this reason a crucial role is played by the functions whose vector field is smooth.

**Definition 17** A function  $\chi \in C^\infty(\mathcal{U}_s, \mathcal{P}_s)$ ,  $\mathcal{U}_s \subset \mathcal{P}_s$  open, is said to be of class  $\text{Gen}_s$ , if the corresponding Hamiltonian vector field  $X_\chi$  is a smooth map from  $\mathcal{U}_s \rightarrow \mathcal{P}_s$ . In this case we will write  $\chi \in \text{Gen}_s$ .

**Proposition 18** Let  $\chi_1 \in \text{Gen}_s$ . If  $\chi_2 \in C^\infty(\mathcal{U}_s, \mathbb{R})$  then  $\{\chi_1, \chi_2\} \in C^\infty(\mathcal{U}_s, \mathbb{R})$ . If  $\chi_2 \in \text{Gen}_s$  then  $\{\chi_1, \chi_2\} \in \text{Gen}_s$ .

**Definition 19** A smooth coordinate transformation  $\mathcal{T} : \mathcal{P}_s \supset \mathcal{U}_s \rightarrow \mathcal{P}_s$  is said to be canonical if for any Hamiltonian function  $H$  one has  $X_{H \circ \mathcal{T}} = \mathcal{T}^* X_H \equiv d\mathcal{T}^{-1} X_H \circ \mathcal{T}$ , i.e. it transforms the Hamilton equations of  $H$  into the Hamilton equations of  $H \circ \mathcal{T}$ .

**Proposition 20** Let  $\chi_1 \in \text{Gen}_s$ , and let  $\Phi_{\chi_1}^t$  be the corresponding time  $t$  flow (which exists by standard theory). Then  $\Phi_{\chi_1}^t$  is a canonical transformation.

## Normal Form for Finite Dimensional Hamiltonian Systems

Consider a system of the form (18), but with finitely many degrees of freedom, namely a system with a Hamiltonian of the form (18) with

$$H_0(p, q) = \sum_{k=1}^n \omega_k \frac{p_k^2 + q_k^2}{2}, \quad \omega_k \in \mathbb{R} \quad (30)$$

and  $H_P$  which is a smooth function having a zero of order at least 3 at the origin.

**Definition 21** A polynomial  $Z$  will be said to be in normal form if  $\{H_0; Z\} \equiv 0$ .

**Theorem 22 (Birkhoff)** For any positive integer  $r \geq 0$ , there exist a neighborhood  $\mathcal{U}^{(r)}$  of the origin and a canonical transformation  $\mathcal{T}_r : \mathbb{R}^{2n} \supset \mathcal{U}^{(r)} \rightarrow \mathbb{R}^{2n}$  which puts the system (18) in Birkhoff Normal Form up to order  $r$ , namely s.t.

$$H^{(r)} := H \circ \mathcal{T}_r = H_0 + Z^{(r)} + \mathcal{R}^{(r)} \quad (31)$$

where  $Z^{(r)}$  is a polynomial of degree  $r+2$  which is in normal form,  $\mathcal{R}^{(r)}$  is small, i.e.

$$|\mathcal{R}^{(r)}(z)| \leq C_r \|z\|^{r+3}, \quad \forall z \in \mathcal{U}^{(r)}; \quad (32)$$

moreover, one has

$$\|z - \mathcal{T}_r(z)\| \leq C'_r \|z\|^2, \quad \forall z \in \mathcal{U}^{(r)}. \quad (33)$$

An inequality identical to (33) is fulfilled by the inverse transformation  $\mathcal{T}_r^{-1}$ .

If the frequencies are nonresonant at order  $r + 2$ , namely if

$$\omega \cdot K \neq 0, \quad \forall K \in \mathbb{Z}^n, \quad 0 < |K| \leq r + 2 \quad (34)$$

the function  $Z^{(r)}$  depends on the actions

$$I_j := \frac{p_j^2 + q_j^2}{2}$$

only.

**Remark 23** If the nonlinearity is analytic and the frequencies are Diophantine, i.e. there exist  $\gamma > 0$  and  $\tau$  such that

$$|\omega \cdot K| \geq \frac{\gamma}{|K|^\tau}, \quad \forall K \in \mathbb{Z}^n - \{0\}, \quad (35)$$

then one can compute the dependence of the constant  $C_r$  (cf. Eq. (32)) on  $r$  and optimize the value of  $r$  as a function of  $\|z\|$ . This allows one to improve (32) and to show that there exists and  $r_{\text{opt}}$  such that (see e.g. (Fassò 1990))

$$|\mathcal{R}^{(r_{\text{opt}})}(z)| \leq C \exp \left( - \frac{c}{\|z\|^{1/(\tau+1)}} \right). \quad (36)$$

In turn, such an estimate is the starting point for the proof of the celebrated Nekhoroshev's theorem (Nekhoroshev 1977).

The idea of the proof is to construct a canonical transformation putting the system in a form which is as simple as possible, namely the normal form. More precisely one constructs a canonical transformation  $\mathcal{T}^{(1)}$  pushing the non normalized part of the Hamiltonian to order four followed by a

transformation  $\mathcal{T}^{(2)}$  pushing it to order five and so on, thus getting  $\mathcal{T}_r = \mathcal{T}^{(1) \circ} \mathcal{T}^{(2) \circ} \dots \circ \mathcal{T}^{(r)}$ . Each of the transformations  $\mathcal{T}^{(j)}$  is constructed as the time one flow of a suitable auxiliary Hamiltonian function say  $\chi_j$  (Lie transform method). It turns out that  $\chi_j$  is determined as the solution of the Homological equation

$$Z_j := \{\chi_j, H_0\} + H^{(j)} \quad (37)$$

where  $H^{(j)}$  is constructed recursively and  $Z_j$  has to be determined together with  $\chi_j$  in such a way that  $\{Z_j, H_0\} = 0$  and (37) holds. In particular  $H^{(1)}$  coincides with the first non vanishing term in the Taylor expansion of  $H_p$ .

The algorithm of solution of the Homological Eq. (37) involves division by the so called small denominators  $i\omega \cdot K$ , where  $K \in \mathbb{Z}^n - \{0\}$ , fulfills  $|K| \leq j + 2$  and  $\omega \cdot K \neq 0$ .

The above construction is more or less explicit: provided one has at disposal enough time, he can explicitly compute  $Z^{(r)}$  up to any given order. In the case of nonresonant frequencies this is not needed if one wants to understand the dynamics over long times. Indeed its features are an easy consequence of the fact that  $Z^{(r)}$  depends on the actions only. A precise statement will be given in the case of PDEs. It has to be noticed that the normal form can be used also as a starting point for the construction of invariant tori through KAM theory. To this end however one has to verify a nondegeneracy condition and this requires the explicit computation of the normal form.

In the resonant case the situation is more complicated, however, it is often enough to compute the first non vanishing term of  $Z^{(r)}$  in order to get relevant information on the dynamics. This usually requires only the ability to compute the function  $Z_1$ , defined by (37) with  $H^{(1)}$  coinciding with the first non vanishing term of the Taylor expansion of  $H_p$ . For a detailed analysis we refer to other sections of the Encyclopedia.

A particular case where one can use a coordinate independent formula for the computation of  $Z_j$  and  $\chi_j$  is the one in which the frequencies are completely resonant.

Assume that there exists  $\nu > 0$  and integer numbers  $\ell_1, \dots, \ell_n$  such that

$$\omega_k = v\ell_k \quad \forall k = 1, \dots, n. \quad (38)$$

**Remark 24** Denote by  $\Psi^t$  the flow of the linear system with Hamiltonian  $H_0$ , then one has

$$\Psi^{t+T} = \Psi^t, \quad T := \frac{2\pi}{v}, \quad t \in \mathbb{R}. \quad (39)$$

Moreover in this case one has  $\omega \cdot K \neq 0 \Rightarrow |\omega \cdot K| \geq v > 0$ , so there are no small denominators.

In this case one has an interesting coordinate independent formula for the solution of the homological Eq. (37).

**Lemma 25** Let  $f$  be smooth function, defined in neighborhood of the origin. Define

$$\begin{aligned} Z(z) &\equiv \langle f \rangle(z) := \frac{1}{T} \int_0^T f(\Psi^t(z)) dt, \\ \chi(z) &:= \frac{1}{T} \int_0^T t[f(\Psi^t(z)) - Z(\Psi^t(z))] dt, \end{aligned} \quad (40)$$

then such quantities fulfill the equation  $\{H_0, \chi\} + f = Z$ .

## Normal Form for Hamiltonian PDEs: General Comments

As anticipated in the introduction there are two problems in order to generalize Birkhoff's theorem to PDEs: (1) the existence of nonsmooth vectorfields and (2) the appearance of small denominators accumulating at zero already at order 3.

There are two reasons why (1) is a problem. The first one is that if the vector field of  $\chi_1$  were not smooth then it would be nontrivial to ensure that it generates a flow, and thus that the normalizing transformation exists. The second related problem is that, if a transformation could be generated, then the Taylor expansion of the transformed Hamiltonian would contain a term of the form  $\{H_1; \chi_1\} = dH_1 X_{\chi_1}$ , which is typically not smooth if  $X_{\chi_1}$  is not smooth. Thus one has to show that the

construction involves only functions which are of class  $\text{Gen}_s$  for some  $s$  (see Definition 17).

The difficulty related to small denominators is the following: In the finite dimensional case,  $\{\omega \cdot K \neq 0, |K| \leq r + 2\}$  implies  $|\omega \cdot K| \geq \gamma > 0$ . Thus division by  $\omega \cdot K$  is a harmless operation in the finite dimensional case. In the infinite dimensional case this is no longer true. For example, when  $\omega_k = \sqrt{k^2 + \mu^2}$  one already has

$$\inf_{0 \neq |K| \leq 3} |\omega \cdot K| = 0.$$

In order to solve such a problem one has to take advantage of a property of the nonlinearity which typically holds in PDEs and is called having localized coefficients. By also assuming a suitable nonresonance property for the frequency vector, one can deduce a normal form theorem identical to Theorem 22. The main difficulty consists in verifying the assumptions of the theorem. We will show how to verify such assumptions by applying this method to some typical examples.

## Normal Form for Resonant Hamiltonian PDEs and Its Consequences

In the case of resonant frequencies and smooth vector field it is possible to obtain a normal form up to an exponentially small remainder.

Consider the system (18) in the phase space  $\mathcal{P}_s$  with some fixed  $s$ . Assume that the frequencies are completely resonant, namely that (38) holds (with  $k \in \mathbb{N}$ ); assume that  $H_P \in \text{Gen}_s$  and that its vector field extends to a complex analytic function in a neighborhood of the origin. Finally assume that  $H_P$  has a zero of order  $n \geq 3$  at the origin. Then we have the following theorem.

**Theorem 26 (Bambusi and Giorgilli 1993; Bambusi and Nekhoroshev 1998)** There exists a neighborhood of the origin  $\mathcal{U}_s \subset \mathcal{P}_s$  and an analytic canonical transformation  $\mathcal{T} : \mathcal{U}_s \rightarrow \mathcal{P}_s$  with the following properties:  $\mathcal{T}$  is close to identity. Namely, it satisfies

$$\|z - \mathcal{T}(z)\|_s \leq C \|z\|_s^{n-1}. \quad (41)$$

$\mathcal{T}$  puts the Hamiltonian in resonant normal form up to an exponentially small remainder, namely one has

$$H^\circ \mathcal{T} = H_0 + \langle H_P \rangle + Z_2 + \mathcal{R} \quad (42)$$

where  $\langle H_P \rangle$  is the average (defined by (40)) of  $H_P$  with respect to the unperturbed flow;  $Z_2$  is in normal form, namely  $\{Z_2; H_0\} \equiv 0$ , and has a zero of order  $2n - 2$  at the origin;  $\mathcal{R}$  is an exponentially small remainder whose vector field is estimated by

$$\|X_{\mathcal{R}}(z)\|_s \leq C \|z\|_s^{n-1} \exp\left(-\frac{C}{\|z\|_s^{n-2}}\right).$$

**Example 27** The nonlinear Schrödinger equation (19). Here one has  $\ell_k = k^2$  and  $v = 1$ . The Sobolev embedding theorems ensure that the vector field of the nonlinearity is analytic if  $f$  is analytic in a neighborhood of the origin. Thus Theorem 26 applies to the NLS. To deduce dynamical consequences it is convenient to explicitly compute  $\langle H_P \rangle$ . Assuming  $f(0) = 0$  and  $f'(0) = 1$  this was done in (Bambusi 1999a) using formula (40) which gives

$$\langle H_P \rangle(z) = \frac{1}{2} \left( \sum_k I_k \right)^2 - \frac{1}{8} \sum_k |I_k|^2 \quad (43)$$

where  $I_k = (p_k^2 + q_k^2)/2$  are the linear actions. Thus one has that  $H_0 + \langle H_P \rangle$  is a function of the actions only, and thus it is an integrable system. It is thus natural to study the system (42) as a perturbation of such an integrable system. This was done in (Bambusi 1999a; Bourgain 2000) obtaining the results we are going to state. For simplicity we will concentrate here on the case of Dirichlet boundary conditions, thus the function  $\psi$  will always be assumed to be skew symmetric with respect to the origin. Define

$$\epsilon_s := \left( \frac{1}{2} \int_{-\pi}^{\pi} |\partial_x^s \psi^{(0)}(x)|^2 dx \right)^{1/2}, \quad (44)$$

i.e., the  $H^s$  norm of the initial datum  $\psi^{(0)}$ ,  $s \geq 0$ , and denote by  $I_k(0)$  the initial value of the linear actions.

**Theorem 28 (Bambusi 1999a)** Fix  $N \geq 1$ , then there exists a constant  $\epsilon_*$ , with the property that, if the initial datum  $\psi^{(0)}$  is such that

$$\epsilon_1 < \epsilon_*, \quad \sum_{k \geq N+1} I_k(0)^2 \leq C \epsilon_1^4 \epsilon_1^{2+1/N}, \quad (45)$$

then along the corresponding solution of (19) one has

$$\sum_{k \geq 1} |I_k(t) - I_k(0)|^2 \leq C' \epsilon_1^{4+1/N} \quad (46)$$

for the times  $t$  fulfilling

$$|t| \leq C'' \exp\left(\frac{\epsilon_*}{\epsilon_1}\right)^{1/N}.$$

This result in particular allows one to control the distance in energy norm of the solution from the torus given by the intersection of the level surfaces of the actions taken at the initial time.

**Theorem 29 (Bourgain 2000)** Fix an arbitrarily large  $r$ , then there exists  $s_r$  such that for any  $s \geq s_r$  there exists  $\epsilon_{*s}$ , such that the following holds true: most of the initial data with  $\epsilon_s < \epsilon_{*s}$  give rise to solutions with

$$\|\psi(t)\|_s \leq C \epsilon_s, \quad \forall |t| \leq \frac{C}{\epsilon_s^r}. \quad (47)$$

For the precise meaning of “most of the initial data,” we refer to the original paper. The result is based on the proof that the considered solutions remain close in the  $H^s$  topology to an infinite dimensional torus. In particular the uniform estimate of the Sobolev norm is relevant for applications to numerical analysis (Cohen et al. 2008; Hairer and Lubich 2006).

**Example 30** Consider the nonlinear wave Eq. (1) with  $\mu = 0$ . Here the frequencies are given by  $|k|$  and thus they are completely resonant. Again the smoothness of the nonlinearity is ensured by Sobolev embedding theorem. In the case of DBC in order to ensure smoothness one has also to assume that the nonlinearity is odd, namely that  $f(-u) = -f(u)$ , then Theorem 26 applies. However in this case the computation of  $\langle H_P \rangle$  is non-trivial. It has been done (see (Palaletti et al. 2001)) in the case of  $f(u) = \pm u^3 + O(u^4)$  and Dirichlet boundary conditions. The result however is that the function  $\langle H_P \rangle$  does not have a particularly simple structure, and thus it is not easy to extract informations on the dynamics.

In order to extract information on the dynamics, consider the simplified system in which the remainder is neglected, namely the system with Hamiltonian

$$H_S := H_0 + \langle H_P \rangle + Z_2. \quad (48)$$

Such a system has two integrals of motion, namely  $H_0$  and  $\langle H_P \rangle + Z_2$ . Let  $\gamma_\epsilon$  be the set of the  $z$ 's at which  $\langle H_P \rangle + Z_2$  is restricted to the surface  $S_\epsilon := \{z : H_0(z) = \epsilon^2\}$  has an extremum, say a maximum. Then  $\gamma_\epsilon$  is an invariant set for the dynamics of  $H_S$ . By the invariance under the flow of  $H_0$ , one has that  $\gamma_\epsilon$  is the union of one dimensional closed curves, but generically it is just a single closed curve. In such a situation it is also a stable periodic orbit of (48) (see (Dell'Antonio 1989)). Actually it is very difficult to compute  $(\langle H_P \rangle + Z_2)|_{S_\epsilon}$ , but a maximum of such a function can be easily constructed by applying the implicit function theorem to a non degenerate maximum of  $\langle H_P \rangle|_{S_\epsilon}$ . The addition of the remainder then modifies the dynamics only after an exponentially long time. We are now going to state the corresponding theorem.

First remark that a critical point of  $\langle H_P \rangle|_{S_1}$  is a solution  $z_a$  of the system

$$\lambda_a A_0 z_a + \nabla \langle H_P \rangle(z_a) = 0, \quad H_0(z_a) = 1 \quad (49)$$

where we used the notations of Remarks 13 and 14. Here  $\lambda_a$  is clearly the Lagrange multiplier. The

closed curve  $\gamma_a := \bigcup_t \Psi^t(z_a)$  is (the trajectory of) a periodic solution of  $H_0 + \langle H_P \rangle$ . Consider now the linear operator  $B_a := d(\nabla \langle H_P \rangle)(z_a)$ .

**Definition 31** The critical point  $z_a$  is said to be non degenerate if the system

$$\lambda_a A_0 h + B_a h = 0, \quad \langle A_0 z_a; h \rangle_0 = 0 \quad (50)$$

has at most one solution.

Under the assumptions of Theorem 32 below it is easy to prove that  $\gamma_a$  is a smooth curve and that its tangent vector  $h_a := \frac{d}{dt} \Psi^t(z_a)|_{t=0}$  is a solution of (50).

**Theorem 32 (Bambusi and Nekhoroshev 1998; Bambusi and Nekhoroshev 2002)** Assume that  $H_P \in Gen_s$  for any  $s$  large enough, assume also that there exists a non degenerate maximum  $z_a$  of  $\langle H_P \rangle|_{S_1}$ , then there exists a constant  $\epsilon_*$ , such that the following holds true: consider a solution  $z(t)$  of the Hamilton equation of (18) with initial datum  $z_0$ ; if there exists  $\epsilon < \epsilon_*$ , such that

$$d_E(\epsilon \gamma_a, z_0) \leq C \epsilon^n, \quad (51)$$

then one has

$$d_E(\epsilon \gamma_a, z(t)) \leq C' \epsilon^n, \quad (52)$$

for all times  $t$  with  $|t| \leq \frac{C}{\epsilon^{n-1}} \exp\left(\frac{\epsilon}{\epsilon_*}\right)^{n-1}$ . Here  $d_E$  is the distance in the energy norm.

Such a theorem does not ensure that there exist periodic orbits of the complete system, but just a family of closed curves with the property that starting close to it one remains close to it for exponentially long times. Some results concerning the existence of true periodic orbits close to such periodic like trajectories can also be proved (see e.g. (Bambusi and Palaletti 2001; Berti and Bolle 2003, 2006; Gentile et al. 2005; Lidskij and Shulman 1988)).

**Example 33** In the paper (Palaletti et al. 2001) it has been proved that the non degeneracy

assumption (50) of Theorem 32 holds for the Eq. (1) with  $f(u) = \pm u^3 +$  higher order terms and Dirichlet boundary conditions. In the case of such an equation an extremum of  $\langle H_P \rangle|_{S_1}$  is given by

$$u(x) = V_m \text{sn}(wx, m), \quad v(x) \equiv 0,$$

with  $V_m$ ,  $w$  and  $m$  suitable constants, and  $\text{sn}$  the Jacobi elliptic sine. Therefore the curve  $\gamma_a$  is the phase space trajectory of the solution of the linear wave equation with such an initial datum. There are no other extrema of  $\langle H_P \rangle|_{S_1}$ . Thus the Theorem 32 ensures that solutions starting close to a rescaling of such a curve remain close to it for very long times.

## Normal Form for Nonresonant Hamiltonian PDEs

### A Statement

We turn now to the nonresonant case. The theory we will present has been developed in (Bambusi 2008; Bambusi et al. 2007b; Grébert 2007), and is closely related to the one of (Bambusi 2003; Bambusi and Grebert 2003; Delort and Szeftel 2006). First we introduce the class of equations to which the theory applies. To this end it is useful to treat the  $p$ 's and the  $q$ 's exactly on an equal footing so we will denote by  $z \equiv (z_k)_{k \in \mathbb{Z}}$ ,  $\mathbb{Z} := \mathbb{Z} - \{0\}$  the set of all the variables, where

$$z_{-k} := p_k, \quad z_k := q_k \quad k \geq 1.$$

Given a polynomial function  $f : \mathcal{P}_\infty \rightarrow \mathbb{R}$  of degree  $r$  one can decompose it as follows

$$f(z) = \sum_{k_1, \dots, k_r} f_{k_1, \dots, k_r} z_{k_1} \dots z_{k_r}. \quad (53)$$

We will assume suitable localization properties for the coefficients  $f_{k_1, \dots, k_r}$  as functions of the indexes  $k_1, \dots, k_r$ .

**Definition 34** Given a multi-index  $k \equiv (k_1, \dots, k_r)$ , let  $(k_{i_1}, k_{i_2}, k_{i_3}, \dots, k_{i_r})$  be a reordering of  $k$  such that

$$|k_{i_1}| \geq |k_{i_2}| \geq |k_{i_3}| \geq \dots \geq |k_{i_r}|.$$

We define  $\mu(k) := |k_{i_3}|$  and

$$S(k) := \mu(k) + \left| |k_{i_1}| - |k_{i_2}| \right|. \quad (54)$$

**Definition 35** Let  $f : \mathcal{P}_\infty \rightarrow \mathbb{R}$  be a polynomial of degree  $r$ . We say that  $f$  has localized coefficients if there exists  $v \in [0, +\infty)$  such that  $\forall N \geq 1$  there exists  $C_N$  such that for any choice of the indexes  $k_1, \dots, k_r$ , the following inequality holds:

$$|f_{k_1, \dots, k_r}| \leq C_N \frac{\mu(k)^{v+N}}{S(k)^N}. \quad (55)$$

**Definition 36** A function  $f \in \text{Gen}_s$  for any  $s$  large enough, is said to have localized coefficients if all the terms of its Taylor expansion have localized coefficients.

Some important properties of functions with localized coefficients are given by Theorem 37.

**Theorem 37** Let  $f : \mathcal{P}_\infty \rightarrow \mathbb{R}$  be a polynomial of degree  $r$  with localized coefficients, then there exists  $s_0$  such that for any  $s \geq s_0$  the vector field  $X_f$  extends to a smooth map from  $\mathcal{P}_s$  to itself; moreover the following estimate holds:

$$\|X_f(z)\| \leq C \|z\|_s \|z\|_{s_0}^{r-2}. \quad (56)$$

In particular it follows that a function with localized coefficients is of class  $\text{Gen}_s$  for any  $s \geq s_0$ .

**Theorem 38** The Poisson Bracket of two functions with localized coefficients has localized coefficients.

In order to develop perturbation theory we also need a quantitative nonresonance condition.

**Definition 39** Fix a positive integer  $r$ . The frequency vector  $\omega$  is said to fulfill the *property (r-NR)* if there exist  $\gamma > 0$ , and  $\alpha \in \mathbb{R}$  such that for any  $N$  large enough one has

$$\left| \sum_{k \geq 1} \omega_k K_k \right| \geq \frac{\gamma}{N^\alpha}, \quad (57)$$

for any  $K \in \mathbb{Z}^\infty$ , fulfilling  $0 \neq |K| := \sum_k |K_k| \leq r + 2$ ,  $\sum_{k > N} |K_k| \leq 2$ .

It is easy to see that under this condition one can solve the homological equation and that, if the known term of the equation has localized coefficients, then the solution also has localized coefficients.

**Theorem 40 (Bambusi et al. 2007b; Bambusi and Grébert 2006)** Fix  $r \geq 1$ , assume that the frequencies fulfill the nonresonance condition ( $r$ -NR); assume that  $H_P$  has localized coefficients. Then there exists a finite  $s_r$ , a neighborhood  $\mathcal{U}_{s_r}^{(r)}$  of the origin in  $\mathcal{P}_{s_r}$ , and a canonical transformation  $\mathcal{T} : \mathcal{U}_{s_r}^{(r)} \rightarrow \mathcal{P}_{s_r}$  which puts the system in normal form up to order  $r + 3$ , namely

$$H^{(r)} := H \circ \mathcal{T} = H_0 + Z^{(r)} + \mathcal{R}^{(r)} \quad (58)$$

where  $Z^{(r)}$  has localized coefficients and is a function of the actions  $I_k$  only;  $\mathcal{R}^{(r)}$  has a small vector field, i.e.

$$\|X_{\mathcal{R}^{(r)}}(z)\|_{s_r} \leq C \|z\|_{s_r}^{r+2}, \quad \forall z \in \mathcal{U}_{s_r}^{(r)}; \quad (59)$$

thus, one has

$$\|z - \mathcal{T}_r(z)\|_{s_r} \leq C \|z\|_{s_r}^2, \quad \forall z \in \mathcal{U}_{s_r}^{(r)}. \quad (60)$$

An inequality identical to (60) is fulfilled by the inverse transformation  $\mathcal{T}^{-1}$ . Finally for any  $s \geq s_r$  there exists a subset  $\mathcal{U}_s^{(r)} \subset \mathcal{U}_{s_r}^{(r)}$  open in  $\mathcal{P}_s$  such that the restriction of the canonical transformation to  $\mathcal{U}_s^{(r)}$  is analytic also as a map from  $\mathcal{P}_s \rightarrow \mathcal{P}_s$  and the inequalities (59) and (60) hold with  $s$  in place of  $s_r$ .

This theorem allows one to give a very precise description of the dynamics.

**Proposition 41** Under the same assumptions of Theorem 37,  $\forall s \geq s_r$  there exists  $\epsilon_{*s}$  such that, if

the initial datum fulfills  $\epsilon := \|z_0\|_s < \epsilon_{*s}$ , then one has

$$\begin{aligned} \|z(t)\|_s &\leq 4\epsilon, \quad \sum_k k^{2s} |I_k(t) - I_k(0)| \\ &\leq C\epsilon^3 \end{aligned} \quad (61)$$

for all the times  $t$  fulfilling  $|t| \leq \epsilon^{-r}$ . Moreover there exists a smooth torus  $\mathbb{T}_0$  such that,  $\forall M \leq r$

$$\begin{aligned} d_s(z(t), \mathbb{T}_0) &\leq C\epsilon^{(M+3)/2}, \quad \text{for } |t| \\ &\leq \frac{1}{\epsilon^{r-M}} \end{aligned} \quad (62)$$

where  $d_s(\cdot, \cdot)$  is the distance in  $\mathcal{P}_s$ .

A generalization to the resonant or partially resonant case is easily obtained and can be found in (Bambusi and Grebert 2003).

### Verification of the Property of Localization of Coefficients

The property of localization of coefficients is quite abstract. We illustrate via a few examples some ways to verify it.

**Example 42** Consider the nonlinear wave Eq. (1) with Neumann boundary conditions on  $[0, \pi]$ . We recall that the corresponding space of functions will be considered as a subset of the space of periodic functions.

Consider the Taylor expansion of the non-linearity, i.e. write  $F(u) = \sum_{r \geq 3} c_r \int_{-\pi}^{\pi} u^r$ . Then one has to prove that the functions  $f_r(u) \equiv \int_{-\pi}^{\pi} u^r$  have localized coefficients. The coefficients  $f_{k_1, \dots, k_r}$  are given by

$$f_{k_1, \dots, k_r} = \int_{-\pi}^{\pi} \cos(k_1 x) \cos(k_2 x) \dots \cos(k_r x) dx. \quad (63)$$

One could compute and estimate such a quantity directly, but it is easier to proceed in a different way: to show that  $f_3$  has localized coefficients and then to use Theorem 38 to show that each  $f_r$  has localized coefficients for any  $r$ . This is the path we will follow. Consider



$$f_{k_1, k_2, k_3} = \int_{-\pi}^{\pi} \cos(k_1 x) \cos(k_2 x) \cos(k_3 x) dx. \quad (64)$$

Since the estimate (55) is symmetric with respect to the indexes, we can assume that they are ordered,  $k_1 \geq k_2 \geq k_3$ , so that (64) =  $\pi \delta_{k_1}^{k_2+k_3} / 2$ ,  $\mu(k) = k_3$ ,  $S(k) = k_3 + k_1 - k_2$  from which one immediately sees that (55) holds with  $v = 0$ . As a consequence one also gets that the function  $g_3 := \int_{-\pi}^{\pi} v u^2$  has localized coefficients. Since  $\{g_3; f_r\} = r f_{r+1}$ , by induction Theorem 38 ensures that  $f_r$  has localized coefficients for any  $r$ .

Often it is impossible to explicitly compute the coefficients  $f_{k_1, \dots, k_r}$ , so one needs a different way to verify the property.

**Example 43** Consider the nonlinear wave equation

$$u_{tt} - u_{xx} + Vu = f(u) \quad (65)$$

with Neumann boundary conditions. Here  $V$  is a smooth, even, periodic potential. The Hamiltonian reduces to the form (18) by introducing the variables  $q_k$  by  $u(x) = \sum_k q_k \varphi_k(x)$  where  $\varphi_k(x)$  are the eigenfunctions of the Sturm Liouville operator  $-\partial_{xx} + V$ , and similarly for  $v$ . In such a case one has

$$f_{k_1, k_2, k_3} = \int_{-\pi}^{\pi} \varphi_{k_1}(x) \varphi_{k_2}(x) \varphi_{k_3}(x) dx. \quad (66)$$

Here the idea is to consider (66) as the matrix element  $L_{k_1, k_2}$  of the operator  $L$  of multiplication by  $\varphi_{k_3}(x)$  on the basis of the eigenfunctions of the operator  $S := -\partial_{xx} + V$ . The key idea is to proceed as follows.

Let  $L$  be a linear operator which maps  $D(S^r)$  into itself for all  $r \geq 0$ , and define the sequence of operators

$$L_N := [S, L_{N-1}], \quad L_0 := L. \quad (67)$$

**Lemma 44** Let  $S$  be as above, then, for any  $N \geq 0$  one has

$$|L_{k_1, k_2}| = |\langle L \varphi_{k_1}; \varphi_{k_2} \rangle| \leq \frac{1}{|\lambda_{k_1} - \lambda_{k_2}|^N} |\langle L_N \varphi_{k_1}; \varphi_{k_2} \rangle| \quad (68)$$

where  $\lambda_{k_j}$  is the eigenvalue of  $S$  corresponding to  $\varphi_{k_j}$ .

Then, in order to conclude the verification of the property of localization of the coefficients, one has just to compute  $L_N$  and to estimate the scalar product product of the r.h.s. All the computations can be found in (Bambusi 2008).

**Example 45** A third example where the verification of the property of localization of coefficients goes almost in the same way is the nonlinear Schrödinger equation

$$i\dot{\psi} = -\psi_{xx} + V\psi + \frac{\partial F(\psi, \bar{\psi})}{\partial \bar{\psi}} \quad (69)$$

with Dirichlet Boundary conditions on  $[0, \pi]$ . Here one has to assume that  $V$  is a smooth even potential and that  $F$  is smooth and fulfills  $F(-\psi, -\bar{\psi}) = F(\psi, \bar{\psi})$  (this is required in order to leave the space of skew symmetric functions invariant, see Remark 12). Here the variables  $(p, q)$  are introduced by

$$\psi = \sum_{k \in \mathbb{Z}} \frac{p_k + iq_k}{\sqrt{2}} \varphi_k, \quad (70)$$

where  $\psi_k$  are the eigenfunctions of  $S$  with Dirichlet boundary conditions. Here the Taylor expansion of the nonlinearity has only even terms. Thus the building block for the proof of the property of localization of coefficients is the operator  $L$  of multiplication by  $\varphi_{k_3} \varphi_{k_4}$ . Then, mutatis mutandis the proof goes as in the previous case.

### Verification of the Nonresonance Property

Finally in order to apply Theorem 40 one has to verify the nonresonance property ( $r$ -NR). As usual in dynamical systems, this is done by tuning the frequencies using parameters. In the case of the nonlinear wave Eq. (1) one can use the mass  $\mu$ .

**Theorem 46 (Bambusi 2003; Bambusi and Grébert 2006; Delort and Szeftel 2004)** There exists a zero measure set  $S \subset \mathbb{R}$  such that, if  $\mu \in \mathbb{R} - S$ , then the frequencies  $\omega_k = \sqrt{k^2 + \mu^2}$ ,  $k \geq 1$  fulfill the condition (r-NR) for any  $r$ .

Thus the Theorem 40 applies to the Eq. (1) with almost any mass.

A similar result holds for the Eq. (65), where the role of the mass is played by the average of the potential.

The situation of the nonlinear Schrödinger is more difficult. Here one can use the Fourier coefficients of the potential as parameters.

Fix  $\sigma > 0$  and, for any positive  $R$  define the space of the potentials, by

$$\mathcal{V}_R := \left\{ V(x) = \sum_{k \geq 1} v_k \cos kx \mid v'_k := v_k R^{-1} e^{\sigma k} \in \left[ -\frac{1}{2}, \frac{1}{2} \right] \text{ for } k \geq 1 \right\} \quad (71)$$

Endow such a space with the product normalized probability measure.

**Theorem 47 (Bambusi and Grébert 2006), see also (Bourgain 1996a)** For any  $r$  there exists a positive  $R$  and a set  $\mathcal{S} \subset \mathcal{V}_R$  such that property (r-NR) holds for any potential  $V \in \mathcal{S}$  and  $|\mathcal{V}_R - \mathcal{S}| = 0$ .

So, provided the potential is chosen in the considered set, Theorem 40 also applies to the Eq. (69).

We point out that the proof of Theorem 46 and of Theorem 47 essentially consists of two steps. First one proves that for most values of the parameters one has

$$\left| \sum_{k \geq 1} \omega_k K_k \right| \geq \frac{\gamma}{N^\alpha}, \quad \forall K \in \mathbb{Z}^\infty, \quad \text{with } 0 \neq |K| := \sum_k |K_k| \leq r + 2. \quad (72)$$

Then one uses the asymptotic of the frequencies, namely  $\omega_k \sim ak^d$  with  $d \geq 1$ , in order to get the result.

## Non Hamiltonian PDEs

In this section we will present some results for the non Hamiltonian case.

It is useful to complexify the phase space. Thus, in this section we will always denote the space of the complex sequences  $\{z_k\}$  whose norm (defined by (25)) is finite by  $\mathcal{P}_s$ .

In the space  $\mathcal{P}_s$  consider a system of differential equations of the form

$$\dot{z}_k = \lambda_k z_k + P_k(z), \quad k \in \mathbb{Z} - \{0\} \quad (73)$$

where  $\lambda_k$  are complex numbers and  $P(z) \equiv \{P_k(z)\}$  has a zero of order at least 2 at the origin. Moreover we will assume  $P$  to be a complex analytic map from a neighborhood of the origin of  $\mathcal{P}_s$  to  $\mathcal{P}_s$ . The quantities  $\lambda_k$  are clearly the eigenvalues of the linear operator describing the linear part of the system, and for this reason they will be called “the eigenvalues”.

**Example 48** A system of the form (18) with  $H_P$  having a vector field which is analytic. The corresponding Hamilton equations have the form (73) with  $\lambda_k = -\lambda_{-k} = i\omega_k$ ,  $k \geq 1$ .

**Example 49** Consider the following nonlinear heat equation with periodic boundary conditions on  $[-\pi, \pi]$ :

$$u_t = u_{xx} - V(x)u + f(u). \quad (74)$$

If  $f$  is analytic then it can be given the form (73) by introducing the basis of the eigenfunctions  $\psi_k$  of the Sturm Liouville operator  $S := -\partial_{xx} + V$ , i.e. denoting  $u = \sum_k z_k \varphi_k$ . In this case the eigenvalues  $\lambda_k$  are the opposite of the periodic eigenvalues of  $S$ . Thus in particular one has  $\lambda_k \in \mathbb{R}$  and  $\lambda_k \sim -k^2$ .

In this context one has to introduce a suitable concept of nonresonance:

**Definition 50** A sequence of eigenvalues is said to be resonant if there exists a sequence of integer numbers  $K_k \geq 0$  and an index  $i$  such that

$$\sum_k \lambda_k K_k - \lambda_i = 0. \quad (75)$$

In the finite dimensional case the most celebrated results concerning systems of the form (73) are the Poincaré theorem, the Poincaré–Dulac theorem and the Siegel theorem (Arnold 1984). The Poincaré theorem is of the form of Birkhoff’s Theorem 22, while the Poincaré–Dulac and Siegel theorems guarantee (under suitable assumptions) that there exists an analytic coordinate transformation reducing the system to its normal form or linear part (no remainder!).

At present there is not a satisfactory extension of the Poincaré–Dulac theorem to PDEs (some partial results have been given in (Foias and Saut 1987)). We now state a known extension of the Siegel theorem to PDEs.

**Theorem 51 (Nikolenko 1986; Zehnder 1978)**

Assume that the eigenvalues fulfill the Diophantine type condition

$$\left| \sum_k \lambda_k K_k - \lambda_i \right| \geq \frac{\gamma}{|K|^\tau}, \quad \forall i, K \quad \text{with } 2 \leq |K|, \quad (76)$$

where  $\gamma > 0$  and  $\tau \in \mathbb{R}$  are suitable parameters; then there exists an analytic coordinate transformation defined in a neighborhood of the origin, such that the system (73) is transformed into its linear part

$$\dot{z}_k = \lambda_k z_k. \quad (77)$$

The main remark concerning this theorem is that the condition (76) is only exceptionally satisfied. If  $\lambda \in \mathbb{C}^n$  then condition (76) is generically satisfied only if  $\tau > (n - 2)/2$ . Nevertheless some examples where such an equation is satisfied are known (Nikolenko 1986).

The formalism of section “Normal Form for Nonresonant Hamiltonian PDEs” can be easily

generalized to the non Hamiltonian case giving rise to a generalization of Poincaré’s theorem, which we are going to state.

Given a polynomial map  $P : \mathcal{P}_\infty \rightarrow \mathcal{P}_{-\infty}$  one can expand it on the canonical basis  $\mathbf{e}_k$  of  $\mathcal{P}_0$  as follows:

$$P(z) = \sum_{k_1, \dots, k_r, i} P_{k_1, \dots, k_r, i}^i z_{k_1} \dots z_{k_r} \mathbf{e}_i, \quad P_{k_1, \dots, k_r, i}^i \in \mathbb{C}. \quad (78)$$

**Definition 52** A polynomial map  $P$  is said to have localized coefficients if there exists  $v \in [0, +\infty)$  such that  $\forall N \geq 1$  there exists  $C_N$  such that for any choice of the indexes  $k_1, \dots, k_r, i$  following the inequality holds:

$$\left| P_{k_1, \dots, k_r, i}^i \right| \leq C_N \frac{\mu(k, i)^{v+N}}{S(k, i)^N}, \quad (79)$$

where  $(k, i) = (k_1, \dots, k_r, i)$ . A map is said to have localized coefficients if for any  $s$  large enough, it is a smooth map from  $\mathcal{P}_s$  to itself and each term of its Taylor expansion has localized coefficients.

**Definition 53** The eigenvalues are said to be strongly nonresonant at order  $r$  if for any  $N$  large enough, any  $K = (K_{k_1}, \dots, K_{k_r})$  and any index  $i$  such that  $|K| \leq r$  and there are at most two of the indexes  $k_1, \dots, k_r, i$  larger than  $N$  the following inequality holds:

$$\left| \sum_k \lambda_k K_k - \lambda_i \right| \geq \frac{\gamma}{N^\alpha}. \quad (80)$$

**Theorem 54** Assume that the nonlinearity has localized coefficients and that the eigenvalues are strongly nonresonant at order  $r$ , then there exists an  $s_r$  and an analytic coordinate transformation  $\mathcal{T}_r : \mathcal{U}_{s_r} \rightarrow \mathcal{P}_{s_r}$  which transforms the system (73) to the form

$$\dot{z}_k = \lambda_k z_k + \mathcal{R}_k(z), \quad (81)$$

where the following inequality holds

$$\|\mathcal{R}(z)\|_{s_r} \leq C \|z\|_{s_r}^r. \quad (82)$$

## Extensions and Related Results

The theory presented here applies to quite general semilinear equations in one space dimension. At present a satisfactory theory applying to quasilinear equations and/or equations in more than one space dimensions is not available. The main difficulty for the extension of the theory to semilinear equations in higher space dimension is related to the nonresonance condition. The general theory can be easily extended to the case where the differences  $\omega_k - \omega_l$  between frequencies accumulate only at a set constituted by isolated points.

**Example 55** Consider the nonlinear wave equation on the  $d$ -dimensional sphere

$$u_{tt} - \Delta_g u + \mu^2 u = f(x, u), \quad x \in S^d \quad (83)$$

with  $\Delta_g$  the Laplace Beltrami operator; the frequencies are given by  $\omega_k = \sqrt{k(k+d-1) + \mu^2}$  and their differences accumulate only at integers. A version of Theorem 40 applicable to (83) was proved in (Bambusi et al. 2007b). As a consequence in particular one has a lower bound on the existence time  $t$  of the small amplitude solutions of the form  $|t| \geq \epsilon^{-r}$ , where  $\epsilon$  is proportional to a high Sobolev norm of the initial datum. An extension to  $u_{tt} - \Delta_g u + Vu = f(x, u)$  is also known.

**Example 56** A similar result was proved in (Bambusi and Grebert 2003) for the nonlinear Schrödinger equation

$$\begin{aligned} -i\dot{\psi} &= -\Delta\psi + V^*\psi \\ &+ f(|\psi|^2)\psi, \quad x \in \mathbb{T}^d, \end{aligned} \quad (84)$$

where the star denotes convolution.

The only general result available at present for quasilinear system is the following theorem.

**Theorem 57 (Bambusi 2005)** Fix  $r \geq 1$ , assume that the frequency vector fulfills condition (72) and that there exists  $d_1$  such that the vector field of  $H_P$  is a smooth map from  $\mathcal{P}_{s+d_1}$  to  $\mathcal{P}_s$  for any  $s$  large enough. Then the same result of Theorem 40 holds, but the functions do not necessarily have localized coefficients and, for any  $s$  large enough **the remainder is estimated by**

$$\|X_{\mathcal{R}^{(r)}}(z)\|_s \leq C \|z\|_{s+d_r}^{r+2}, \quad \forall z \in \mathcal{U}_s^{(r)}; \quad (85)$$

where  $d_r$  is a large positive number.

The estimate (85) shows that the remainder is small only when considered as an operator extracting a lot of derivatives. In particular it is non trivial to use such a theorem in order to extract information on the dynamics. Following the approach of (Bambusi 2005; Bambusi et al. 2002) this can be done using the normal form to construct approximate solutions and suitable versions of the Gronwall lemma to compare it with solutions of the true system. This however allows to control the dynamics only over times of the order of  $\epsilon^{-1}$ ,  $\epsilon$  being again a measure of the size of the initial datum. Such a theory has been applied to quasilinear wave equations in (Bambusi 2005) and to the Fermi Pasta Ulam problem in (Bambusi and Ponno 2006). Among the large number of papers containing related results we recall (Foias and Saut 1987; Krol 1989; Pals 1996; Stroucken and Verhulst 1987). A stronger result for the nonlinear wave equation valid over times of order  $\epsilon^{-2}$  can be found in (Delort and Szeftel 2004).

## Future Directions

Future directions of research include both purely theoretical aspects and applications to other sciences.

From a purely theoretical point of view, the most important open problems pertain to the validity of normal form theory for equations in which the nonlinearity involves derivatives, and for general equations in more than one space dimension.

These results would be particularly important since the kind of equations appearing in most domains of physics are quasilinear and higher dimensional.

Concerning applications, we would like to describe a few of them which would be of interest.

- Water wave problem. The problem of description of the free surface of a fluid has been shown to fit in the scheme of Hamiltonian dynamics (Zakharov 1968). Normal form theory could allow one to extract the relevant informations on the dynamics in different situations (Craig 1996; Dyachenko and Zakharov 1994), ranging from the theory of Tsunamis (Craig 2006) to the theory of fluid interface, which is relevant e.g. to the construction of oil platforms (Craig et al. 2005).
- Quantum mechanics. A Bose condensate is known to be well described by the Gross Pitaevskii equation. When the potential is confining, such an equation is of the form (18). Normal form theory has already been used for some preliminary results (Bambusi and Sacchetti 2007), but a systematic investigation could lead to interesting new results.
- Electromagnetic theory and magneto hydrodynamics. The equations have a Hamiltonian form; normal form theory could help to describe some instability arising in plasmas.
- Elastodynamics. Here one of the main theoretical open problems is the stability of equilibria which are a minimum of the energy. The problem is that in higher dimensions the conservation of energy does not ensure enough smoothness of the solution to ensure stability. Such a problem is of the same kind as the one solved in (Bambusi et al. 2007b) when dealing with the existence times of the nonlinear wave equation.

## Bibliography

Arnold V (1984) Chapitres supplémentaires de la théorie des équations différentielles ordinaires. Mir, Moscow  
 Bambusi D (1999a) Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equation. *Math Z* 130:345–387

Bambusi D (1999b) On the Darboux theorem for weak symplectic manifolds. *Proc Am Math Soc* 127(11): 3383–3391  
 Bambusi D (2003) Birkhoff normal form for some nonlinear PDEs. *Commun Math Phys* 234:253–283  
 Bambusi D (2005) Galerkin averaging method and Poincaré normal form for some quasilinear PDEs. *Ann Sci Norm Super Pisa Cl Sci* 4(5):669–702  
 Bambusi D (2008) A Birkhoff normal form theorem for some semilinear PDEs. In: Craig W (ed) *Hamiltonian dynamical systems and applications*. Springer  
 Bambusi D, Giorgilli A (1993) Exponential stability of states close to resonance in infinite-dimensional Hamiltonian systems. *J Stat Phys* 71(3–4):569–606  
 Bambusi D, Grebert B (2003) Forme normale pour NLS en dimension quelconque. *C R Acad Sci Paris* 337: 409–414  
 Bambusi D, Grébert B (2006) Birkhoff normal form for partial differential equations with tame modulus. *Duke Math J* 135(3):507–567  
 Bambusi D, Nekhoroshev NN (1998) A property of exponential stability in the nonlinear wave equation close to main linear mode. *Phys D* 122:73–104  
 Bambusi D, Nekhoroshev NN (2002) Long time stability in perturbations of completely resonant PDE's, symmetry and perturbation theory. *Acta Appl Math* 70(1–3):1–22  
 Bambusi D, Paleari S (2001) Families of periodic orbits for resonant PDE's. *J Nonlinear Sci* 11:69–87  
 Bambusi D, Ponno A (2006) On metastability in FPU. *Commun Math Phys* 264(2):539–561  
 Bambusi D, Sacchetti A (2007) Exponential times in the one-dimensional Gross–Pitaevskii equation with multiple well potential. *Commun Math Phys* 234(2):136  
 Bambusi D, Carati A, Ponno A (2002) The nonlinear Schrödinger equation as a resonant normal form. *DCDS-B* 2:109–128  
 Bambusi D, Carati A, Penati T (2007a) On the relevance of boundary conditions for the FPU paradox. Preprint *Inst Lombardo Accad Sci Lett Rend A* (to appear)  
 Bambusi D, Delort JM, Grébert B, Szeftel J (2007b) Almost global existence for Hamiltonian semi-linear Klein–Gordon equations with small Cauchy data on Zoll manifolds. *Commun Pure Appl Math* 60(11): 1665–1690  
 Bambusi D, Muraro D, Penati T (2008) Numerical studies on boundary effects on the FPU paradox. *Phys Lett A* 372(12):2039–2042  
 Berti M, Bolle P (2003) Periodic solutions of nonlinear wave equations with general nonlinearities. *Commun Math Phys* 243:315–328  
 Berti M, Bolle P (2006) Cantor families of periodic solutions for completely resonant nonlinear wave equations. *Duke Math J* 134(2):359–419  
 Bourgain J (1996a) Construction of approximative and almost-periodic solutions of perturbed linear Schrödinger and wave equations. *Geom Funct Anal* 6:201–230  
 Bourgain J (1996b) On the growth in time of higher Sobolev norms of smooth solutions of Hamiltonian PDE. *Int Math Res Not* 6:277–304

- Bourgain J (1997) On growth in time of Sobolev norms of smooth solutions of nonlinear Schrödinger equations in  $R^D$ . *J Anal Math* 72:299–310
- Bourgain J (1998) Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equation. *Ann Math* 148:363–439
- Bourgain J (2000) On diffusion in high-dimensional Hamiltonian systems and PDE. *J Anal Math* 80:1–35
- Bourgain J (2005a) Green's function estimates for lattice Schrödinger operators and applications. In: *Annals of mathematics studies*, vol 158. Princeton University Press, Princeton
- Bourgain J (2005b) On invariant tori of full dimension for 1D periodic NLS. *J Funct Anal* 229(1):62–94
- Bourgain J, Kaloshin V (2005) On diffusion in high-dimensional Hamiltonian systems. *J Funct Anal* 229(1):1–61
- Chernoff PR, Marsden JE (1974) Properties of infinite dimensional Hamiltonian systems. In: *Lecture notes in mathematics*, vol 425. Springer, Berlin
- Cohen D, Hairer E, Lubich C (2008) Long-time analysis of nonlinearly perturbed wave equations via modulated Fourier expansions. *Arch Ration Mech Anal* 187(2):341–368
- Craig W (1996) Birkhoff normal forms for water waves. In: *Mathematical problems in the theory of water waves* (Luminy, 1995), *Contemp Math*, vol 200. Amer Math Soc, Providence, pp 57–74
- Craig W (2006) Surface water waves and tsunamis. *J Dynam Differ Equ* 18(3):525–549
- Craig W, Wayne CE (1993) Newton's method and periodic solutions of nonlinear wave equations. *Commun Pure Appl Math* 46:1409–1498
- Craig W, Guyenne P, Kalisch H (2005) Hamiltonian long-wave expansions for free surfaces and interfaces. *Commun Pure Appl Math* 58(12):1587–1641
- Dell'Antonio GF (1989) Fine tuning of resonances and periodic solutions of Hamiltonian systems near equilibrium. *Commun Math Phys* 120(4):529–546
- Delort JM, Szeftel J (2004) Long-time existence for small data nonlinear Klein–Gordon equations on tori and spheres. *Int Math Res Not* 37:1897–1966
- Delort J-M, Szeftel J (2006) Long-time existence for semi-linear Klein–Gordon equations with small Cauchy data on Zoll manifolds. *Am J Math* 128(5):1187–1218
- Dyachenko AI, Zakharov VE (1994) Is free-surface hydrodynamics an integrable system? *Phys Lett A* 190:144–148
- Eliasson HL, Kuksin SB (2006) KAM for non-linear Schrödinger equation. *Ann Math. Preprint* (to appear)
- Fassò F (1990) Lie series method for vector fields and Hamiltonian perturbation theory. *Z Angew Math Phys* 41(6):843–864
- Foias C, Saut JC (1987) Linearization and normal form of the Navier–Stokes equations with potential forces. *Ann Inst H Poincaré Anal Non Linéaire* 4:1–47
- Gentile G, Mastropietro V, Procesi M (2005) Periodic solutions for completely resonant nonlinear wave equations with Dirichlet boundary conditions. *Commun Math Phys* 256(2):437–490
- Grébert B (2007) Birkhoff normal form and Hamiltonian PDES. *Partial differential equations and applications*, 1–46 Sémin Congr, 15 Soc Math France, Paris
- Hairer E, Lubich C (2006) Conservation of energy, momentum and actions in numerical discretizations of nonlinear wave equations
- Kappeler T, Pöschel J (2003) *KdV & KAM*. Springer, Berlin
- Klainerman S (1983) On almost global solutions to quasi-linear wave equations in three space dimensions. *Commun Pure Appl Math* 36:325–344
- Krol MS (1989) On Galerkin-averaging method for weakly nonlinear wave equations. *Math Methods Appl Sci* 11:649–664
- Kuksin SB (1987) Hamiltonian perturbations of infinite-dimensional linear systems with an imaginary spectrum. *Funct Anal Appl* 21:192–205
- Kuksin SB (1993) Nearly integrable infinite-dimensional Hamiltonian systems. Springer, Berlin
- Kuksin SB, Pöschel J (1996) Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. *Ann Math* 143:149–179
- Lidskij BV, Shulman EI (1988) Periodic solutions of the equation  $u_{tt} - u_{xx} + u^3 = 0$ . *Funct Anal Appl* 22:332–333
- Marsden J (1972) Darboux's theorem fails for weak symplectic forms. *Proc Am Math Soc* 32:590–592
- Nekhoroshev NN (1977) Exponential estimate of the stability of near integrable Hamiltonian systems. *Russ Math Surv* 32(6):1–65
- Nikolenko NV (1986) The method of Poincaré normal form in problems of integrability of equations of evolution type. *Russ Math Surv* 41:63–114
- Paléari S, Bambusi D, Cacciatori S (2001) Normal form and exponential stability for some nonlinear string equations. *ZAMP* 52:1033–1052
- Pals H (1996) The Galerkin-averaging method for the Klein–Gordon equation in two space dimensions. *Nonlinear Anal TMA* 27:841–856
- Pöschel J (2002) On the construction of almost-periodic solutions for a nonlinear Schrödinger equation. *Ergodic Theory Dynam Syst* 22:1–22
- Soffer A, Weinstein MI (1999) Resonances, radiation damping and instability in Hamiltonian nonlinear wave equations. *Invent Math* 136(1):9–74
- Stroucken ACJ, Verhulst F (1987) The Galerkin-averaging method for nonlinear, undamped continuous systems. *Math Methods Appl Sci* 335:520–549
- Wayne CE (1990) Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory. *Commun Math Phys* 127:479–528
- Weinstein A (1969) Symplectic structures on Banach manifolds. *Bull Am Math Soc* 75:1040–1041
- Zakharov VE (1968) Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Appl Mech Tech Phys* 2:190–194
- Zehnder E (1978) C L Siegel's linearization theorem in infinite dimensions. *manuscripta math* 23:363–371



# Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite Dimensional Systems

Luigi Chierchia and Michela Procesi  
Dipartimento di Matematica e Fisica, Università  
“Roma Tre”, Rome, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Finite Dimensional KAM Theory  
Infinite Dimensional KAM Theory  
Future Directions  
Appendix A: The Classical Implicit Function Theorem  
Appendix B: Complementary Notes  
Bibliography

## Glossary

**Action-angles variables** A particular set of variables  $(y, x) = ((y_1, \dots, y_d), (x_1, \dots, x_d))$ ,  $x_i$  (“angles”) defined modulus  $2\pi$ , particularly suited to describe the general behavior of a finite dimensional integrable system.

**Complex symplectic variables** The identification of the real symplectic space  $(\mathbb{R}^{2d}, dp \wedge dq)$  with the complex space  $\mathbb{C}^d$ : one sets  $z_j = (q_j + ip_j)/\sqrt{2}$  for  $j = 1, \dots, d$ , in this way the symplectic two-form is the imaginary part of the hermitian product.

**Fast convergent (Newton) method** Super-exponential algorithms, mimicking Newton’s method of tangents, used to solve differential problems involving small divisors.

**Hamiltonian dynamics** The dynamics generated by a Hamiltonian differential equation on a symplectic space/manifold (in the finite

dimensional case, an even-dimensional manifold endowed with a symplectic structure).

**Hamiltonian System** A time reversible, conservative (without dissipation or expansion) dynamical system, which generalizes classical mechanical systems (solutions of Newton’s equation  $m_i \ddot{x}_i = f_i(x)$ , with  $1 \leq i \leq d$  and  $f = (f_1, \dots, f_d)$  a conservative force field); they are described by the flow of differential equations (i.e., the time  $t$  map associating to an initial condition, the solution of the initial value problem at time  $t$ ) on a symplectic space/manifold.

**Integrable Hamiltonian systems** Very special class of Hamiltonian systems, whose flow can be “explicitly computed” for all initial data and typically is described through a linear flow on a (in)finite dimensional torus.

**Invariant tori** Manifolds diffeomorphic to tori invariant for the flow of a differential equation (especially, of Hamiltonian differential equations); establishing the existence of tori invariant for Hamiltonian flows is the main object of KAM theory.

**KAM** Acronym from the names of Kolmogorov (Andrey Nikolaevich Kolmogorov, 1903–1987), Arnold (Vladimir Igorevich Arnold, 1937–2010) and Moser (Jürgen K. Moser, 1928–1999), whose results, in the 1950s and 1960s, in Hamiltonian dynamics, gave rise to the theory presented in this article.

**Nearly-integrable Hamiltonian systems** Hamiltonian systems which are small perturbations of an integrable systems and which, in general, exhibits a much richer dynamics than the integrable limit. Nevertheless, finite dimensional KAM theory asserts that, under suitable assumptions, the majority (in measure sense) of the initial data of a nearly integrable system behaves as in the integrable limit.

**Quasi-periodic motions** Trajectories (solutions of a system of differential equations), which are conjugate to linear flow on tori  $x \in \mathbb{T}^d \mapsto$



$x + \omega t$  with  $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$  called frequency vector.

**Small divisors/denominators** Arbitrary small combinations of the form  $\omega \cdot k := \sum_{j=1}^d \omega_j k_j$  with  $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$  a real vector and  $k \in \mathbb{Z}^d$  an integer vector different from zero; these combinations arise in the denominators of certain expansions appearing in the perturbation theory of Hamiltonian systems, making (when  $d > 1$ ) convergent arguments very delicate. Physically, small divisors are related to “resonances,” which are a typical feature of conservative systems.

**Stability** The property of orbits of having certain properties similar to a reference limit; more specifically, in the context of KAM theory, stability is normally referred to the property of action variables of staying close to their initial values.

**Symplectic structure** A mathematical structure (a differentiable, nondegenerate, closed 2-form) apt to describe, in an abstract setting, the main geometrical features of conservative differential equations arising in mechanics.

## Definition of the Subject

KAM theory is a mathematical, quantitative theory which has as primary object the persistence, under small (Hamiltonian) perturbations, of quasi-periodic trajectories of integrable Hamiltonian systems. Quasi-periodic motions may be described through the linear flow  $x \in \mathbb{T}^d \rightarrow x + \omega t \in \mathbb{T}^d$  where  $\mathbb{T}^d$  denotes the standard  $d$ -dimensional torus (see section “[Introduction](#)” below),  $t$  is time, and  $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$  is the set of frequencies of the trajectory (if  $d = 1$ ,  $2\pi/\omega$  is the *period* of the motion).

In finite dimensional integrable systems with bounded motions, the typical trajectory is indeed quasi-periodic and KAM theory is apt to describe the behavior of “most” initial data. In general, this is not the case in infinite dimensional systems and PDEs. Still, the search for periodic and quasi-periodic solutions is obviously an interesting and challenging task.

## Introduction

The main motivation for KAM theory is related to stability questions arising in celestial mechanics which were addressed by astronomers and mathematicians such as Kepler, Newton, Lagrange, Liouville, Delaunay, Weierstrass, and, from a more modern point of view, Poincaré, Birkhoff, Siegel, etc.

The major breakthrough, in this context, was due to Kolmogorov in 1954, followed by the fundamental work of Arnold and Moser in the early 1960s, who were able to overcome the formidable technical problem related to the appearance, in perturbative formulae, of arbitrarily small divisors<sup>1</sup>. Small divisors make impossible the use of classical analytical tools (such as the standard Implicit Function Theorem, fixed point theorems) and could be controlled only through a “fast convergent method” of Newton-type<sup>2</sup>, which allowed, in view of the super-exponential rate of convergence, to counterbalance the divergences introduced by small divisors.

KAM theory was extended to the context of Hamiltonian PDEs starting from the early 1990s by Kuksin, Wayne, Pöschel, with the purpose of proving the existence and linear stability of *small*-quasi-periodic solutions for semi-linear PDEs with Dirichlet boundary conditions. Although there is no general theory available, the KAM approach has been successively developed in order to cover also examples of PDEs on tori and compact Lie groups, quasi-linear and fully nonlinear PDEs on the circle, and PDEs on the line with a coercive potential.

Actually, the main bulk of KAM theory is a set of *techniques* based, as mentioned, on fast convergent methods and solving various questions in Hamiltonian (or generalizations of Hamiltonian) dynamics. There are excellent reviews of KAM theory – especially Sect. 6.3 of Arnold et al. (2006) and Sevryuk (2003) – which should complement the reading of this article, whose main objective is not to review but rather to explain the main fundamental ideas of KAM theory. To do this, we re-examine, in modern language, the main ideas introduced, respectively, by the founders of KAM theory, namely, Kolmogorov (in section “[Kolmogorov Theorem](#)”), Arnold

(in section “[Arnold’s Scheme](#)”), and Moser (section “[The Differentiable Case: Moser’s Theorem](#)”).

While the tools and techniques in finite dimensions are by now quite well understood, the situation in infinite dimensions is, as can be expected, significantly more complicated and there are many fundamental open issues, such as the “general” behavior of a “nearly integrable” system. Therefore, in discussing the finite dimensional case, we shall try to give a quite complete and quantitative description of results and, especially, the techniques used in order to obtain them. On the other hand, in infinite dimension, we mainly focus on specific examples, trying to convey the main ideas and the similarities and differences with the finite dimensional case.

A set of technical notes (such as notes 17, 18, 19, 21, 24, 26, 29, 30, 31, 34, 39), which the reader not particularly interested in technical mathematical arguments may skip, are collected in [Appendix B](#) and complete the mathematical exposition. [Appendix B](#) includes also several other complementary notes, which contain either standard material or further references or side comments.

In section “[Future Directions](#),” we briefly and informally describe a few developments and applications of KAM theory: this section is by no means exhaustive and is meant to give a non-technical, short introduction to some of the most important (in our opinion) extensions of the original contributions.

### Finite Dimensional Context

In the finite dimensional setting, we will be concerned with Hamiltonian flows on the symplectic manifold  $(\mathcal{M}, dy \wedge dx)$ ; for general information, see, e.g., Arnold (1974) or Sect. 1.3 of Arnold et al. (2006). Notation, main definitions, and properties are listed in the following items.

- (a)  $\mathcal{M} := B \times \mathbb{T}^d$  with  $d \geq 2$  (the case  $d = 1$  is trivial for the questions addressed in this entry);  $B$  is an open, connected, bounded set in  $\mathbb{R}^d$ ;  $\mathbb{T}^d := \mathbb{R}^d / (2\pi\mathbb{Z}^d)$  is the standard flat  $d$ -dimensional torus with periods<sup>3</sup>  $2\pi$ ;

- (b)  $dy \wedge dx := \sum_{i=1}^d dy_i \wedge dx_i$ , ( $y \in B, x \in \mathbb{T}^d$ ) is the standard symplectic form<sup>4</sup>;
- (c) Given a real-analytic (or smooth) function  $H: \mathcal{M} \rightarrow \mathbb{R}$ , the *Hamiltonian flow governed by  $H$*  is the one-parameter family of diffeomorphisms  $\phi_H^t: \mathcal{M} \rightarrow \mathcal{M}$ , which to  $z \in \mathcal{M}$  associates the solution at time  $t$  of the differential equation<sup>5</sup>

$$\dot{z} = J_{2d} \nabla H(z), \quad z(0) = z, \quad (1)$$

where  $\dot{z} = \frac{dz}{dt}$ ,  $J_{2d}$  is the standard symplectic  $(2d \times 2d)$ -matrix  $J_{2d} = \begin{pmatrix} 0 & -1_d \\ 1_d & 0 \end{pmatrix}$ ,  $1_d$  denotes the unit  $(d \times d)$ -matrix and  $0$  denotes a  $(d \times d)$  block of zeros, and  $\nabla$  denotes gradient; in the symplectic coordinates  $(y, x) \in B \times \mathbb{T}^d$ , Eq. (1) reads

$$\begin{cases} \dot{y} = -H_x(y, x) \\ \dot{x} = H_y(y, x) \end{cases}, \quad \begin{cases} y(0) = y \\ x(0) = x \end{cases} \quad (2)$$

Clearly, the flow  $\phi_H^t$  is defined until  $y(t)$  reaches eventually the border of  $B$ .

Equations (1) or (2) are called the *Hamilton’s equations* with *Hamiltonian*  $H$ ; usually, the symplectic (or “conjugate”) variables  $(y, x)$  are called *action-angles* variables<sup>6</sup>; the number  $d$  (= half of the dimension of the phase space) is also referred to as “the number of degrees of freedom<sup>7</sup>.”

The Hamiltonian  $H$  is constant over trajectories  $\phi_H^t(z)$ , as it follows immediately by differentiating  $t \rightarrow H(\phi_H^t(z))$ . The constant value  $E = H(\phi_H^t(z))$  is called the energy of the trajectory  $\phi_H^t(z)$ .

Hamilton equations are left invariant by *symplectic* (or “canonical”) change of variables, i.e., by diffeomorphisms of  $\mathcal{M}$  which preserve the 2-form  $dy \wedge dx$ ; i.e., if  $\phi: (\eta, \xi) \in \mathcal{M} \rightarrow (y, x) \in \mathcal{M}$  is a diffeomorphism such that  $d\eta \wedge d\xi = dy \wedge dx$ , then

$$\phi^{-1} \circ \phi_H^t \circ \phi = \phi_{H \circ \phi}^t. \quad (3)$$

An equivalent condition for a map  $\phi$  to be symplectic is that its Jacobian  $\phi'$  is a *symplectic matrix*, i.e.,

$$\phi'^T J_{2d} \phi' = J_{2d} \quad (4)$$

where  $J_{2d}$  is the standard symplectic matrix introduced above and the superscript  $T$  denotes matrix transpose.

By a (generalization of a) theorem of Liouville, the Hamiltonian flow is symplectic, i.e., the map  $(y, x) \rightarrow (\eta, \xi) = \phi_H^t(y, x)$  is symplectic for any  $H$  and any  $t$ ; see Corollary 1.8, Arnold et al. (2006).

A classical way of producing symplectic transformations is by means of *generating functions*. For example, if  $g(\eta, x)$  is a smooth function of  $2d$  variables with

$$\det \frac{\partial^2 g}{\partial \eta \partial x} \neq 0,$$

then, by the IFT (Implicit Function Theorem; see Kolmogorov and Fomin (1999) or Appendix A below), the map  $\phi: (y, x) \rightarrow (\eta, \xi)$  defined implicitly by the relations

$$y = \frac{\partial g}{\partial x}, \quad \xi = \frac{\partial g}{\partial \eta},$$

yields a local symplectic diffeomorphism; in such a case,  $g$  is called the generating function of the transformation  $\phi$ ; the function  $\eta \cdot x$  is the generating function of the identity map.

For general information about symplectic changes of coordinates, generating functions, and, in general, about symplectic structures, we refer the reader to Arnold (1974) or Arnold et al. (2006).

- (d) A solution  $z(t) = (y(t), x(t))$  of (2) is a *maximal quasi-periodic solution* with frequency vector  $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$  if  $\omega$  is a rationally independent vector, i.e.,

$$\begin{aligned} \exists n \in \mathbb{Z}^d \text{ s.t. } \quad \omega \cdot n &:= \sum_{i=1}^d \omega_i n_i = 0 \\ \Rightarrow \quad n &= 0, \end{aligned} \quad (5)$$

and if there exist smooth (periodic) functions  $v, u: \mathbb{T}^d \rightarrow \mathbb{R}^d$  such that<sup>8</sup>

$$\begin{cases} y(t) = v(\omega t) \\ x(t) = \omega t + u(\omega t) \end{cases} \quad (6)$$

is a solution of (2).

- (e) Let  $\omega, u$  and  $v$  be as in the preceding item and let  $U$  and  $\phi$  denote, respectively, the maps

$$\begin{cases} U: \theta \in \mathbb{T}^d \rightarrow U(\theta) := \theta + u(\theta) \in \mathbb{T}^d \\ \phi: \theta \in \mathbb{T}^d \rightarrow \phi(\theta) := (v(\theta), U(\theta)) \in \mathcal{M} \end{cases}$$

If  $U$  is a smooth diffeomorphism of  $\mathbb{T}^d$  (so that, in particular<sup>9</sup>  $\det U_\theta \neq 0$ ) then  $\phi$  is an embedding of  $\mathbb{T}^d$  into  $\mathcal{M}$  and the set

$$\mathcal{T}_\omega = \mathcal{T}_\omega^d := \phi(\mathbb{T}^d) \quad (7)$$

is an embedded  $d$ -torus invariant for  $\phi_H^t$  and on which the motion is conjugated to the linear (Kronecker) flow  $\theta \rightarrow \theta + \omega t$ , i.e.,

$$\phi^{-1} \circ \phi_H^t \circ \phi(\theta) = \theta + \omega t, \quad \forall \theta \in \mathbb{T}^d. \quad (8)$$

Furthermore, the invariant torus  $\mathcal{T}_\omega$  is a graph over  $\mathbb{T}^d$  and is *Lagrangian*, i.e., the restriction of the symplectic form  $dy \wedge dx$  on  $\mathcal{T}_\omega$  vanishes<sup>10</sup>.

- (f) In KAM theory, a major rôle is played by the numerical properties of the frequencies  $\omega$ . A typical assumption is that  $\omega$  is a *Diophantine vector*:  $\omega \in \mathbb{R}^d$  is called Diophantine or  $(\kappa, \tau)$ -Diophantine if, for some constants  $0 < \kappa \leq \min_i |\omega_i|$  and  $\tau \geq d - 1$ , it verifies the following inequalities:

$$|\omega \cdot n| \geq \frac{\kappa}{|n|^\tau}, \quad \forall n \in \mathbb{Z}^d \setminus \{0\}, \quad (9)$$

(normally, for integer vectors  $n$ ,  $|n|$  denotes  $|n_1| + \dots + |n_d|$ , but other norms may as well be used). We shall refer to  $\kappa$  and  $\tau$  as the Diophantine constants of  $\omega$ . The set of Diophantine numbers in  $\mathbb{R}^d$  with constants  $\kappa$  and  $\tau$  will be denoted by  $\mathcal{D}_{\kappa, \tau}$  or  $\mathcal{D}_{\kappa, \tau}^d$ , while the union over all  $\kappa > 0$  of  $\mathcal{D}_{\kappa, \tau}$  will be denoted by  $\mathcal{D}_\tau = \mathcal{D}_\tau^d$ . Basic facts about these sets are<sup>11</sup>: if  $\tau < d - 1$  then  $\mathcal{D}_\tau^d = \emptyset$  if  $\tau > d - 1$  then the Lebesgue measure of  $\mathbb{R}^d \setminus \mathcal{D}_\tau$  is zero; if  $\tau = d - 1$ , the Lebesgue measure of  $\mathcal{D}_\tau$  is

zero, but its intersection with any open set has the cardinality of  $\mathbb{R}$ . The union over all  $\tau \geq d - 1$  of  $\mathcal{D}_\tau^d$  will be denoted by  $\mathcal{D}^d$ .

- (g) The tori  $\mathcal{T}_\omega$  defined in (e) with  $\omega \in \mathcal{D}^d$  will be called *maximal KAM tori*.
- (h) A Hamiltonian function  $(\eta, \xi) \in \mathcal{M} \rightarrow H(\eta, \xi)$  having a maximal KAM torus (or, more in general, a maximal invariant torus as in (e) with  $\omega$  rationally independent)  $\mathcal{T}_\omega$  can be put into the form.<sup>12</sup>

$$K := E + \omega \cdot y + Q(y, x) \quad \text{with} \quad (10)$$

$$\partial_y^\alpha Q(0, x) = 0, \quad \forall \alpha \in \mathbb{N}^d, |\alpha| \leq 1;$$

compare, e.g., Sect. 1 of Salamon (2004). A Hamiltonian in the form (10) is said to be in *Kolmogorov normal form*.

If

$$\det \langle \partial_y^2 Q(0, \cdot) \rangle \neq 0, \quad (11)$$

(where the brackets denote average over  $\mathbb{T}^d$  and  $\partial_y^2$  the Hessian with respect to the  $y$ -variables) we shall say that the Kolmogorov normal form  $K$  in (10) is *nondegenerate*; similarly, we shall say that the KAM torus  $\mathcal{T}_\omega$  is nondegenerate if it admits a nondegenerate Kolmogorov normal.

- (i) Quasi-periodic solutions with  $1 \leq n < d$  frequencies, i.e., solutions of (2) of the form

$$\begin{cases} y(t) = v(\omega t) \\ x(t) = U(\omega t) \end{cases} \quad (12)$$

where  $v: \mathbb{T}^n \rightarrow \mathbb{R}^d$ ,  $U: \mathbb{T}^n \rightarrow \mathbb{T}^d$  are smooth functions,  $\omega \in \mathbb{R}^n$  is a rationally independent  $n$ -vector. Also in this case, if the map  $U$  is a diffeomorphism onto its image, the set

$$\mathcal{T}_\omega^n := \{(y, x) \in \mathcal{M} : y = v(\theta), x = U(\theta), \quad \theta \in \mathbb{T}^n\} \quad (13)$$

defines an invariant  $n$ -torus on which the flow  $\phi_H^t$  acts by the linear translation  $\theta \rightarrow \theta + \omega t$ . Such tori are normally referred to as *lower dimensional tori*.

**Remark 1** (i) A classical theorem by H. Weyl says that the flow

$$\theta \in \mathbb{T}^d \rightarrow \theta + \omega t \in \mathbb{T}^d, \quad t \in \mathbb{R}$$

is dense (ergodic) in  $\mathbb{T}^d$  if and only if  $\omega \in \mathbb{R}^d$  is rationally independent (compare Arnold et al. (2006), Theorem 5.4 of Katok and Hasselblatt (1995), Sect. 1.4). Thus, trajectories on KAM tori fill them densely (i.e., pass in any neighborhood of any point).

(ii) In view of the preceding remark, it is easy to see that if  $\omega$  is rationally independent,  $(v(t), x(t))$  in (6) is a solution of (2) if and only if the functions  $v$  and  $u$  satisfy the following quasi-linear system of PDE's on  $\mathbb{T}^d$ :

$$\begin{cases} D_\omega v = -H_x(v(\theta), \theta + u(\theta)) \\ \omega + D_\omega u = H_y(v(\theta), \theta + u(\theta)) \end{cases} \quad (14)$$

where  $D_\omega$  denotes the directional derivative  $\sum_{i=1}^d \omega_i \frac{\partial}{\partial \theta_i}$ .

(iii) Probably, the main motivation for studying quasi-periodic solutions of Hamiltonian systems on  $\mathbb{R}^d \times \mathbb{T}^d$  comes from perturbation theory for *nearly integrable* Hamiltonian systems: a completely integrable system may be described by a Hamiltonian system on  $\mathcal{M} := B(y_0, r) \times \mathbb{T}^d \subset \mathbb{R}^d \times \mathbb{T}^d$  with Hamiltonian  $H = K(y)$  (compare Theorem 5.8, Arnold et al. (2006)); here  $B(y_0, r)$  denotes the open ball  $\{y \in \mathbb{R}^d : |y - y_0| < r\}$  centered at  $y_0 \in \mathbb{R}^d$ ; we shall also denote by  $D(y_0, r)$  the complex ball in  $\mathbb{C}^d$  of radius  $r$  centered in  $y_0 \in \mathbb{C}^d$ . In such a case, the Hamiltonian flow is simply

$$\begin{aligned} \phi_K^t(y, x) &= (y, x + \omega(y)t), \\ \omega(y) &:= K_y(y) := \frac{\partial K}{\partial y}(y). \end{aligned} \quad (15)$$

Thus, if the frequency map  $y \in B \rightarrow \omega(y)$  is a diffeomorphism (which is guaranteed if  $\det K_{yy}(y_0) \neq 0$ , for some  $x_0 \in B$  and  $B$  is small enough), in view of (f), for almost all initial data, the trajectories (15) belong to maximal KAM tori  $\{y\} \times \mathbb{T}^d$  with  $\omega(y) \in \mathcal{D}^d$ .

The main content of (classical) KAM theory, in our language, is that, *if the frequency map  $\omega$  is a diffeomorphism, KAM tori persist under small*

perturbations of  $K$ ; compare Remark 7–(iv) below.

The study of the dynamics generated by the flow of a one-parameter family of Hamiltonians of the form

$$K(y) + \varepsilon P(y, x; \varepsilon), \quad 0 < \varepsilon \ll 1, \quad (16)$$

was called by H. Poincaré *le problème général de la dynamique*, to which he dedicated a large part of his monumental *Méthodes Nouvelles de la Mécanique Céleste* (Poincaré).

### Infinite Dimensional Context

In the infinite dimensional setting, we will be concerned with Hamiltonian flows on a scale of Banach or Hilbert spaces; for a more detailed presentation we refer the reader to Kuksin (2000, 2004), Grébert and Kappeler (2014), Kappeler and Pöschel (2003); for properties of analytic functions on Hilbert spaces, see Pöschel and Trubowitz (1987).

- (A) A symplectic structure on scale of real Hilbert spaces  $\mathcal{H}_s$ ,  $(\cdot, \cdot)_s$  is defined by an antisymmetric morphism  $J: \mathcal{H}_s \rightarrow \mathcal{H}_s +_d$  of order  $d \geq 0$  so that the symplectic two form defined by  $\omega(u, v) := (u, J^{-1}v)_0$  is a skew-symmetric bilinear form.
- (B) Given a complex Hilbert space  $H$  with a Hermitian product  $\langle \cdot, \cdot \rangle$ , its realification is a real symplectic Hilbert space with scalar product and symplectic form given by

$$(u, v) = 2\operatorname{Re}\langle u, v \rangle, \quad \omega(u, v) = 2\operatorname{Im}\langle u, v \rangle.$$

- (C) Given a real-analytic (or smooth) function  $H: O_s \subseteq \mathcal{H}_s \rightarrow \mathbb{R}$ ,  $O_s$  open, the *Hamiltonian flow governed by  $H$*  is, if the equation is at least locally well posed, the one-parameter family of diffeomorphisms  $\phi_H^t: \mathcal{H}_s \rightarrow \mathcal{H}_s$ , which to  $z \in \mathcal{H}_s$  associates the solution at time  $t$  of the differential equation

$$\dot{z} = J\nabla H(z), \quad z(0) = z \in \mathcal{H}_s, \quad (17)$$

where  $J$  is the symplectic morphism and  $\nabla H$  is identified through the bilinear product  $(\cdot)_0$  namely

$$dH[\cdot] = (\nabla H, \cdot)_0.$$

Note that in general in the infinite dimensional case the fact that  $H$  is smooth does not guarantee that the equation is even locally well posed. As in the finite dimensional counterpart, the Hamiltonian  $H$  is constant over trajectories  $\phi_H^t(z)$ , and the constant value  $E = H(\phi_H^t(z))$  is called the energy of the trajectory  $\phi_H^t(z)$ .

- (D) A Hamiltonian  $H$  whose Hamiltonian vector field is an analytic map  $\mathcal{H}_s \rightarrow \mathcal{H}_s$  is called a *regular* Hamiltonian; in this case the Hamiltonian flow is at least locally well posed.
- (E) Hamilton equations are left invariant by *symplectic* (or “canonical”) change of variables, i.e., by diffeomorphisms of  $\mathcal{H}_s$  which preserve the 2-form  $\omega$ . A classical way to generate such changes of coordinates is as the time-one flow of an auxiliary Hamiltonian function, say  $S$ . We recall that given a smooth Hamiltonian  $S: O_s \rightarrow \mathbb{R}$  if the corresponding Hamilton equation

$$w_\tau = J\nabla S(w), \quad w|_{\tau=0} = z$$

is well posed for  $\tau \leq 1$  then the flow  $\phi_H^1(z)$  defines a symplectic change of variables.

- (F) A solution  $z(t)$  of (17) is *quasi-periodic* with frequency vector  $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$  if  $\omega$  is a rationally independent vector, see (5), and there exists an embedding  $U: \mathbb{T}^d \rightarrow \mathcal{H}_s$  such that  $z(t) = U(\omega t)$  is a solution of (17).
- (G) A partial differential equation, supplemented by some boundary conditions, is called a Hamiltonian partial differential equation, or an HPDE, if under a suitable choice of a symplectic Hilbert scale, domain, and Hamiltonian, it can be written in the form (17).
- (H) A nonlinear PDE is called *fully nonlinear* if the highest order derivatives appear with degree higher than one, and it is called *quasi-linear* if the highest order derivatives appear with degree one both in the linear and in the nonlinear terms of the equation. It is called *semi-linear* if the linear term contains derivatives of higher order with respect to the nonlinear terms.

(I) Integrable PDEs is a fascinating, deep, and interesting field by its own, and it has been widely studied starting from the 1960s with a variety of methods (formal algebraic methods, algebraic geometry, inverse spectral methods, ...). For the connection of infinite integrable systems and KAM methods, see, e.g., Kappeler and Pöschel (2003) and Kuksin (2004).

(ii) In fact, *the dependence upon  $\varepsilon$  is analytic* and therefore the torus  $\mathcal{T}_{\omega, \varepsilon}$  is an analytic deformation of the unperturbed torus  $\mathcal{T}_{\omega, 0}$  (which is invariant for  $K$ ); see Remark 7-(iii) below.

(iii) Actually, Kolmogorov not only stated the above result but gave also a precise outline of its proof, which is based on a *fast convergent “Newton” scheme*, as we shall see below.

The map  $\phi_*$  is obtained as

$$\phi_* = \lim_{j \rightarrow \infty} \phi_1 \circ \dots \circ \phi_j,$$

where the  $\phi_j$ 's are ( $\varepsilon$ -dependent) symplectic transformations of  $\mathcal{M}$  closer and closer to the identity. It is enough to describe the construction of  $\phi_1$ ;  $\phi_2$  is then obtained by replacing  $H_0 := H$  with  $H_1 = H \circ \phi_1$  and so on.

We proceed to analyze the scheme of Kolmogorov's proof, which will be divided into three main steps.

#### Step 1: Kolmogorov Transformation

The map  $\phi_1$  is close to the identity and it is generated by

$$g(y', x) := y' \cdot x + \varepsilon(b \cdot x + s(x) + y' \cdot a(x))$$

where  $s$  and  $a$  are (respectively, scalar- and vector-valued) real-analytic functions on  $\mathbb{T}^d$  with zero average and  $b \in \mathbb{R}^d$ : setting

$$\begin{aligned} \beta_0 &= \beta_0(x) := b + s_x, \\ A &= A(x) := a_x \quad \text{and} \quad \beta = \beta(y', x) := \beta_0 + Ay', \end{aligned} \quad (19)$$

$(s_x = \partial_x s = (s_{x_1}, \dots, s_{x_d}))$  and  $a_x$  denotes the matrix  $(a_x)_{ij} := \frac{\partial a_i}{\partial x_j}$ )  $\phi_1$  is implicitly defined by

$$\begin{cases} y = y' + \varepsilon \beta(y', x) := y' + \varepsilon(\beta_0(x) + A(x)y') \\ x' = x + \varepsilon a(x). \end{cases} \quad (20)$$

Thus, for  $\varepsilon$  small,  $x \in \mathbb{T}^d \rightarrow x + \varepsilon a(x) \in \mathbb{T}^d$  defines a diffeomorphism of  $\mathbb{T}^d$  with inverse

$$x = \varphi(x') := x' + \varepsilon \alpha(x'; \varepsilon), \quad (21)$$

for a suitable real-analytic function  $\alpha$ , and  $\phi_1$  is explicitly given by

## Finite Dimensional KAM Theory

### Kolmogorov Theorem

In the 1954 International Congress of Mathematicians, in Amsterdam, A.N. Kolmogorov announced the following fundamental (for the terminology, see (f), (g) and (h) above).

**Theorem 1 (Kolmogorov (1954))** *Consider a one-parameter family of real-analytic Hamiltonian functions on  $\mathcal{M} := B(0, r) \times \mathbb{T}^d$  given by*

$$H := K + \varepsilon P, \quad (\varepsilon \in \mathbb{R}), \quad (18)$$

where: (i)  $K$  is a nondegenerate Kolmogorov normal form; (ii)  $\omega \in \mathcal{D}^d$  is Diophantine. Then, there exists  $\varepsilon_0 > 0$  and for any  $|\varepsilon| \leq \varepsilon_0$  a real-analytic symplectic transformation  $\phi_*: \mathcal{M}^* := B(0, r_*) \times \mathbb{T}^d \rightarrow \mathcal{M}$ , for some  $0 < r_* < r$ , putting  $H$  in nondegenerate Kolmogorov normal form,  $H \circ \phi_* = K_*$ , with  $K_* := E_* + \omega \cdot y' + Q_*(y', x')$ . Furthermore<sup>13</sup>,  $\|\phi_* - \text{id}\|_{C^1(\mathcal{M}_*)}$ ,  $|E_* - E|$ , and  $\|Q_* - Q\|_{C^1(\mathcal{M}_*)}$  are small with  $\varepsilon$ .

**Remark 2** (i) From Theorem 1, it follows that the torus

$$\mathcal{T}_{\omega, \varepsilon} := \phi_*(0, \mathbb{T}^d)$$

is a maximal nondegenerate KAM torus for  $H$  and the  $H$ -flow on  $\mathcal{T}_{\omega, \varepsilon}$  is analytically conjugated (by  $\phi_*$ ) to the translation  $x' \rightarrow x' + \omega t$  with the same frequency vector of  $\mathcal{T}_{\omega, 0} := \{0\} \times \mathbb{T}^d$ , while the energy of  $\mathcal{T}_{\omega, \varepsilon}$ , namely,  $E_*$ , is in general different from the energy  $E$  of  $\mathcal{T}_{\omega, 0}$ . The idea of keeping fixed the frequency is a key idea introduced by Kolmogorov, and its importance will be made clear in the analysis of the proof.



$$\begin{aligned} \phi_1 : (y', x') \\ \rightarrow \begin{cases} y = y' + \varepsilon\beta(y', \varphi(y', \varphi(x'))) \\ x = \varphi(x'). \end{cases} \end{aligned} \quad (22)$$

**Remark 3** (i) Kolmogorov transformation  $\phi_1$  is actually the composition of two “elementary” symplectic transformations:  $\phi_1 = \phi_1^{(1)} \circ \phi_1^{(2)}$  where  $\phi_1^{(2)} : (y', x') \rightarrow (\eta, \xi)$  is the symplectic lift of the  $\mathbb{T}^d$ -diffeomorphism given by  $x' = \xi + \varepsilon a(\xi)$  (i.e.,  $\phi_1^{(2)}$  is the symplectic map generated by  $y' \cdot \xi + \varepsilon y' \cdot a(\xi)$ ), while  $\phi_1^{(1)} : (\eta, \xi) \rightarrow (y, x)$  is the angle-dependent action translation generated by  $\eta \cdot x + \varepsilon(b \cdot x + s(x))$ ;  $\phi_1^{(2)}$  acts in the “angle direction” and will be needed to straighten out the flow up to order  $O(\varepsilon^2)$ , while  $\phi_1^{(1)}$  acts in the “action direction” and will be needed to keep the frequency of the torus fixed.

(ii) The inverse of  $\phi_1$  has the form

$$(y, x) \rightarrow \begin{cases} y' = M(x)y + c(x) \\ x' = \varphi(x) \end{cases} \quad (23)$$

with  $M$  a  $(d \times d)$ -invertible matrix and  $\varphi$  a diffeomorphism of  $\mathbb{T}^d$  (in the present case  $M = (1_d + \varepsilon A(x))^{-1} = 1_d + O(\varepsilon)$  and  $\varphi = \text{id} + \varepsilon a$ ) and it is easy to see that the symplectic diffeomorphisms of the form (23) form a subgroup of the symplectic diffeomorphisms, which we shall call *the group of Kolmogorov transformation*.

**Determination of Kolmogorov transformation.** Following Kolmogorov, we now try to determine  $b$ ,  $s$ , and  $a$  so that the “new Hamiltonian” (better: “the Hamiltonian in the new symplectic variables”) takes the form

$$H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1, \quad (24)$$

with  $K_1$  in the Kolmogorov normal form

$$K_1 = E_1 + \omega \cdot y' + Q_1(y', x'), \quad (Q_1 = O(|y'|^2)). \quad (25)$$

To proceed we insert  $y = y' + \varepsilon\beta(y', x)$  into  $H$  and, after some elementary algebra and using Taylor formula, we find<sup>14</sup>

$$\begin{aligned} H(y' + \varepsilon\beta, x) &= E + \omega \cdot y' + Q'(y', x) \\ &\quad + \varepsilon F(y', x) + \varepsilon^2 P'(y', x) \end{aligned} \quad (26)$$

where, letting

$$\begin{aligned} Q^{(1)} &:= [Q_y(y', x) - Q_{yy}(0, x)y'] \cdot \beta_0 \\ &= \frac{1}{2} \int_0^1 Q_{yyy}(ty', x)y' \cdot y' \cdot \beta_0 dt \\ Q^{(2)} &:= P(y', x) - P(0, x) - P_y(0, x)y' \\ &= \frac{1}{2} \int_0^1 P_{yy}(ty', x)y' \cdot y' dt \\ P^{(1)} &:= \frac{1}{\varepsilon^2} [Q(y' + \varepsilon\beta, x) - Q(y', x) - \varepsilon Q_y(y', x) \cdot \beta] \\ &= \frac{1}{2} \int_0^1 Q_{yy}(y' + t\varepsilon\beta, x)\beta \cdot \beta dt P^{(2)} \\ &:= \frac{1}{\varepsilon} [P(y' + \varepsilon\beta, x) - P(y', x)] \\ &= \int_0^1 P_y(y' + t\varepsilon\beta, x) \cdot \beta dt, \end{aligned} \quad (27)$$

(recall that  $Q_y(0, x) = 0$ ) and denoting the  $\omega$ -directional derivative

$$D_\omega := \sum_{j=1}^d \omega_j \frac{\partial}{\partial x_j}$$

one sees that  $Q' = Q'(y', x)$ ,  $F = F(y', x)$  and  $P' = P'(y', x)$  are given by, respectively

$$\begin{aligned} Q' &:= Q(y', x) + \varepsilon \tilde{Q}(y', x), \\ \tilde{Q}(y', x) &:= Q_y(y', x) \cdot (a_x y') + Q^{(1)} + Q^{(2)} \\ F &:= b \cdot \omega + D_\omega s + D_\omega a \cdot y' \\ &\quad + Q_{yy}(0, x)y' \cdot \beta_0 + P(0, x) + P_y(0, x) \cdot y' \\ P' &:= P^{(1)} + P^{(2)}, \end{aligned} \quad (28)$$

where  $D_\omega a$  is the vector function with  $k$ th-entry  $\sum_{j=1}^d \omega_j \frac{\partial a_k}{\partial x_j}$ ;  $D_\omega a \cdot y' = \omega \cdot (a_x y') = \sum_{j,k=1}^d \omega_j \frac{\partial a_k}{\partial x_j} y'_k$ ; recall, also, that  $Q = O(|y'|^2)$  so that  $Q_y = O(y)$  and  $Q' = O(|y'|^2)$ .

Notice that, as intermediate step, we are considering  $H$  as a function of mixed variables  $y'$  and  $x$  (and this causes no problem, as it will be clear along the proof).

Thus, recalling that  $x$  is related to  $x'$  by the ( $y'$ -independent) diffeomorphism  $x = x' + \varepsilon a(x'; \varepsilon)$



in (22), we see that in order to achieve relations (24)-(25), we have to determine  $b$ ,  $s$ , and  $a$  so that

$$F(y', x) = \text{const.} \quad (29)$$

**Remark 4** (i)  $F$  is a first degree polynomial in  $y'$  so that (29) is equivalent to

$$\begin{cases} b \cdot \omega + D_\omega s + P(0, x) = \text{const}, \\ D_\omega a + Q_{yy}(0, x)\beta_0 + P_y(0, x) = 0. \end{cases} \quad (30)$$

Indeed, the second equation is necessary to keep the torus frequency fixed and equal to  $\omega$ , which, as we shall see in more detail later, is a key ingredient introduced by Kolmogorov. (ii) In solving (29) or (30), we shall encounter differential equations of the form

$$D_\omega u = f, \quad (31)$$

for some given function  $f$  real-analytic on  $\mathbb{T}^d$ . Taking the average over  $\mathbb{T}^d$  shows that  $\langle f \rangle = 0$ , and we see that Eq. (31) can be solved only if  $f$  has vanishing mean value

$$\langle f \rangle = f_0 = 0;$$

in such a case, expanding in Fourier series<sup>15</sup>, one sees that (31) is equivalent to

$$\sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} i\omega \cdot n u_n e^{in \cdot x} = \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} f_n e^{in \cdot x}, \quad (32)$$

so that the solutions of (31) are given by

$$u = u_0 + \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{f_n}{i\omega \cdot n} e^{in \cdot x}, \quad (33)$$

for an arbitrary  $u_0$ . Recall that for a continuous function  $f$  over  $\mathbb{T}^d$  to be analytic is necessary and sufficient that its Fourier coefficients  $f_n$  decay exponentially fast in  $n$ , i.e., that there exist positive constants  $M$  and  $\sigma$  such that

$$|f_n| \leq M e^{-\sigma|n|}, \quad \forall n. \quad (34)$$

Now, since  $\omega \in \mathcal{D}_{\kappa, \tau}$  one has that

$$\frac{1}{|\omega \cdot n|} \leq \frac{|n|^\tau}{\kappa} \quad (35)$$

and one sees that if  $f$  is analytic so is  $u$  in (33) (although the decay constants of  $u$  will be different to those of  $f$ ; see below).

Summarizing, *if  $f$  is real-analytic on  $\mathbb{T}^d$  and has vanishing mean value  $f_0$ , then there exists a unique real-analytic solution of (31) with vanishing mean value, which is given by*

$$D_\omega^{-1} f := \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{f_n}{i\omega \cdot n} e^{in \cdot x}; \quad (36)$$

all other solutions of (31) are obtained by adding an arbitrary constant to  $D_\omega^{-1} f$  as in (33) with  $u_0$  arbitrary.

Taking the average of the first relation in (30), we may determine the value of the constant denoted const, namely,

$$\text{const} = b \cdot \omega + P_0(0) := b \cdot \omega + \langle P(0, \cdot) \rangle. \quad (37)$$

Thus, by (ii) of Remark 4, we see that

$$\begin{aligned} s &= D_\omega^{-1} (P(0, x) - P_0(0)) \\ &= \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{P_n(0)}{i\omega \cdot n} e^{in \cdot x}, \end{aligned} \quad (38)$$

where  $P_n(0)$  denote the Fourier coefficients of  $x \mapsto P(0, x)$ ; indeed  $s$  is determined only up to a constant by the relation in (30) but *we select the natural zero-average solution*. Thus,  $s$  has been completely determined.

To solve the second (vector) equation in (30), we first have to require that the l.h.s. (left hand side) has vanishing mean value, i.e., recalling that  $\beta_0 = b + s_x$  (see (19)), we must have

$$\langle Q_{yy}(0, \cdot) \rangle b + \langle Q_{yy}(0, \cdot) s_x \rangle + \langle P_y(0, \cdot) \rangle = 0. \quad (39)$$

In view of (11), this relation is equivalent to

$$b = -\left\langle Q_{yy}(0, \cdot) \right\rangle^{-1} \left( \left\langle Q_{yy}(0, \cdot) s_x \right\rangle + \left\langle P_y(0, \cdot) \right\rangle \right), \quad (40)$$

which determines uniquely  $b$ . Thus,  $\beta_0$  is completely determined, the l.h.s. of the second equation in (30) has zero average and the unique zero-average solution (again zero-average of  $a$  is required as a normalization condition) is given by

$$a = -D_\omega^{-1} \left( Q_{yy}(0, x) \beta_0 + P_y(0, x) \right). \quad (41)$$

Finally, if  $\varphi(x') = x' + \varepsilon \alpha(x'; \varepsilon)$  is the inverse diffeomorphism of  $x \rightarrow x + \varepsilon a(x)$  (compare (21)), then by Taylor's formula,

$$Q(y'(\varphi(x'))) = Q(y', x') + \varepsilon \int_0^1 Q_x(y', x' + \varepsilon \alpha t) \cdot \alpha dt.$$

In conclusion, we have proved.

**Proposition 1** *If  $\phi_1$  is defined in (20)–(19) with  $s$ ,  $b$ , and  $a$  given in (38), (40), and (41), respectively, then (24) holds with*

$$\begin{aligned} E_1 &:= E + \varepsilon \tilde{E}, \\ \tilde{E} &:= b \cdot \omega + P_0(0) Q := \int_0^1 Q_x(y', x' + t \varepsilon \alpha) \cdot \alpha dt \\ &\quad + Q'(y', \varphi(x')), P_1(y', x') := P'(y', \varphi(x')) \end{aligned} \quad (42)$$

with  $Q'$  and  $P'$  defined in (27), (28) and  $\varphi$  in (21).

**Remark 5** The main technical problem is now transparent: because of the appearance of the *small divisors*  $\omega \cdot n$  (which may become arbitrarily small), the solution  $D_\omega^{-1} f$  is *less regular* than  $f$  so that the approximation scheme cannot work on a fixed function space. To overcome this fundamental problem – which even Poincaré was unable to solve notwithstanding his enormous efforts (see, e.g., [Poincaré](#)) – three ingredients are necessary:

- (i) To set up a Newton scheme: this step has just been performed and it has been summarized in the above Proposition 1; such schemes have the following fundamental

advantages: they are “quadratic” and furthermore, after one step one has reproduced the initial situation (i.e., the form of  $H_1$  in (24) has the same properties of  $H_0$ ). It is important to notice that the new perturbation  $\varepsilon^2 P_1$  is proportional to the *square*  $\varepsilon$ ; thus, if one could iterate, at the  $j$ th step, would find

$$H_j = H_{j-1} \circ \phi_j = K_j + \varepsilon^{2j} P_j. \quad (43)$$

The appearance of the exponential of the exponential of  $\varepsilon$  justifies the term “super-converge” used, sometimes, in connection with Newton schemes.

- (ii) One needs to introduce a *scale of Banach function spaces*  $\{\mathcal{B}_\xi; \xi > 0\}$  with the property that  $\mathcal{B}_{\xi'} \subset \mathcal{B}_\xi$  when  $\xi < \xi'$ : the generating functions  $\phi_j$  will belong to  $\mathcal{B}_{\xi_j}$  for a suitable decreasing sequence  $\xi_j$ .
- (iii) One needs to control the small divisors at each step, and this is granted by Kolmogorov's idea of keeping *fixed* the frequency in the normal form so that one can systematically use the Diophantine estimate (9).

Kolmogorov in his paper explained very neatly steps (i) and (iii) but did not provide the details for step (ii); at this regard he added: “*Only the use of condition (9) for proving the convergence of the recursions,  $\phi_j$ , to the analytic limit for the recursion  $\phi_*$  is somewhat more subtle.*” In the next paragraph, we shall introduce classical Banach spaces and discuss the needed straightforward estimates.

## Step 2: Estimates

For  $\xi \leq 1$ , we denote by  $\mathcal{B}_\xi$  the space of functions  $f: B(0, \xi) \times \mathbb{T}^d \rightarrow \mathbb{R}$  analytic on

$$W_\xi := D(0, \xi) \times \mathbb{T}_\xi^d, \quad (44)$$

where

$$\begin{aligned} D(0, \xi) &:= \{y \in \mathbb{C}^d : |y| < \xi\} \\ \text{and } \mathbb{T}_\xi^d &:= \{x \in \mathbb{C}^d : |\operatorname{Im} x_j| < \xi\} / (2\pi \mathbb{Z}^d) \end{aligned} \quad (45)$$

with finite sup-norm

$$\|f\|_{\xi} := \sup_{D(0, \xi) \times \mathbb{T}_{\xi}^d} |f|, \quad (46)$$

(in other words,  $\mathbb{T}_{\xi}^d$  denotes the complex points  $x$  with real parts  $\operatorname{Re} x_j$  defined modulus  $2\pi$  and imaginary part  $\operatorname{Im} x_j$  with absolute value less than  $\xi$ ).

The following properties are elementary:

- (P1)  $\mathcal{B}_{\xi}$  equipped with the  $\|\cdot\|_{\xi}$  norm is a Banach space  
(P2)  $\mathcal{B}_{\xi'} \subset \mathcal{B}_{\xi}$  when  $\xi < \xi'$  and  $\|f\|_{\xi} \leq \|f\|_{\xi'}$  for any  $f \in \mathcal{B}_{\xi'}$   
(P3) if  $f \in \mathcal{B}_{\xi}$ , and  $f_n(y)$  denotes the  $n$ -Fourier coefficient of the periodic function  $x \rightarrow f(y, x)$ , then

$$|f_n(y)| \leq \|f\|_{\xi} e^{-|n|\xi}, \quad \forall n \in \mathbb{Z}^d, \quad (47)$$

Another elementary property, which together with (P3) may be found in any book of complex variables (e.g., Ahlfors (1978)), is the following “Cauchy estimate” (which is based on Cauchy integral formula):

- (P4) let  $f \in \mathcal{B}_{\xi}$  and let  $p \in \mathbb{N}$  then there exists a constant  $B_p = B_p(d) \geq 1$  such that, for any multiindex  $(\alpha, \beta) \in \mathbb{N}^d \times \mathbb{N}^d$  with  $|\alpha| + |\beta| \leq p$  (as above for integer vectors  $\alpha$ ,  $|\alpha| = \sum_j |\alpha_j|$ ) and for any  $0 \leq \xi' < \xi$  one has

$$\left\| \partial_y^{\alpha} \partial_x^{\beta} f \right\|_{\xi'} \leq B_p \|f\|_{\xi} (\xi - \xi')^{-(|\alpha| + |\beta|)}. \quad (48)$$

Finally, we shall need estimate on  $D_{\omega}^{-1}f$ , i.e., on solutions of (31):

- (P5) Assume that  $x \rightarrow f(x) \in \mathcal{B}_{\xi}$  has zero average; assume that  $\omega \in \mathcal{D}_{\kappa, \tau}$  (recall point (f) of section “Introduction”), and let  $p \in \mathbb{N}$ . Then, there exist constants  $\bar{B}_p = \bar{B}_p(d, \tau) \geq 1$  and  $k_p = k_p(d, \tau) \geq 1$  such that, for any multiindex  $\beta$  in  $\mathbb{N}^d$  with  $|\beta| \leq p$  and for any  $0 \leq \xi' < \xi$  one has

$$\left\| \partial_x^{\beta} D_{\omega}^{-1} f \right\|_{\xi'} \leq \bar{B}_p \frac{\|f\|_{\xi}}{\kappa} (\xi - \xi')^{-k_p}. \quad (49)$$

**Remark 6** (i) A proof of (49) is easily obtained observing that by (36) and (47), calling  $\delta := \xi - \xi'$ , one has

$$\begin{aligned} \left\| \partial_x^{\beta} D_{\omega}^{-1} f \right\|_{\xi'} &\leq \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{|n|^{|\beta|} |f_n|}{|\omega \cdot n|} e^{\xi' |n|} \\ &\leq \|f\|_{\xi} \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} \frac{|n|^{|\beta| + \tau}}{\kappa} e^{-\delta |n|} \\ &= \frac{\|f\|_{\xi}}{\kappa} \delta^{-(|\beta| + \tau + d)} \sum_{\substack{n \in \mathbb{Z}^d \\ n \neq 0}} [\delta |n|]^{|\beta| + \tau} e^{-\delta |n|} \delta^d \\ &\leq \text{const} \frac{\|f\|_{\xi}}{\kappa} (\xi - \xi')^{-(|\beta| + \tau + d)}, \end{aligned}$$

where last estimate comes from approximating the sum with the Riemann integral

$$\int_{\mathbb{R}^d} |y|^{|\beta| + \tau} e^{-|y|} dy.$$

More surprising (and much more subtle) is that (49) holds with  $k_p = |\beta| + \tau$ ; such estimate has been obtained by Rüssmann (1975, 1976). For other explicit estimates, see, e.g., Celletti and Chierchia (1988) or Celletti and Chierchia (1995).

(ii) If  $|\beta| > 0$ , it is not necessary to assume that  $\langle f \rangle = 0$ .

(iii) Other norms may be used (and, sometimes, are more useful); for example, rather popular are Fourier norms

$$\|f\|'_{\xi} := \sum_{n \in \mathbb{Z}^d} |f_n| e^{\xi |n|}; \quad (50)$$

see, e.g., Celletti and Chierchia (2007) and references therein.

By the hypotheses of Theorem 1, it follows that there exist  $0 < \xi \leq 1$ ,  $\kappa > 0$  and  $\tau \geq d - 1$  such that  $H \in \mathcal{B}_{\xi}$  and  $\omega \in \mathcal{D}_{\kappa, \tau}$ . Denote

$$T := \left\langle Q_{yy}(0, \cdot) \right\rangle^{-1}, \quad M := \|P\|_{\xi}. \quad (51)$$

and let  $C > 1$  be a constant such that<sup>16</sup>

$$|E|, |\omega|, \|Q\|_{\xi}, \|T\| < C \quad (52)$$

(i.e., each term on the l.h.s. is bounded by the r.h.s.); finally, fix

$$0 < \delta < \xi \quad \text{and define} \quad \bar{\xi} := \xi - \frac{2}{3}\delta, \quad \xi' := \xi - \delta. \quad (53)$$

The parameter  $\xi'$  will be the size of the domain of analyticity of the new symplectic variables  $(y', x')$ , domain on which we shall bound the

Hamiltonian  $H_1 = H \circ \phi_1$ , while  $\bar{\xi}$  refers to an intermediate domain where we shall bound various functions of  $y'$  and  $x$ .

By (P4) and (P5), it follows that there exist constants  $\bar{c} = \bar{c}(d, \tau, \kappa) > 1$ ,  $\bar{\mu} \in \mathbb{Z}_+$  and  $\bar{\nu} = \bar{\nu}(d, \tau) > 1$  such that<sup>17</sup>

$$\|s_x\|_{\bar{\xi}}, \|b\|, |\tilde{E}|, \|a\|_{\bar{\xi}}, \|a_x\|_{\bar{\xi}}, \|\beta_0\|_{\bar{\xi}}, \|\beta\|_{\bar{\xi}}, \quad \|Q'\|_{\bar{\xi}}, \left\| \partial_{y'}^2 Q'(0, \cdot) \right\|_0 \leq \bar{c} C^{\bar{\mu}} \delta^{-\bar{\nu}} M =: \bar{L}, \|P'\|_{\bar{\xi}} \leq \bar{c} C^{\bar{\mu}} \delta^{-\bar{\nu}} M^2 =: \bar{L}M. \quad (54)$$

The estimate in (54) allows to construct, for  $\varepsilon$  small enough, the symplectic transformation  $\phi_1$ , whose main properties are collected in the following.

**Lemma 1** *If  $|\varepsilon| \leq \varepsilon_0$  and  $\varepsilon_0$  satisfies*

$$\varepsilon_0 \bar{L} \leq \frac{\delta}{3}, \quad (55)$$

*then the map the map  $\psi_\varepsilon(x) := x + \varepsilon a(x)$  has an analytic inverse  $\varphi(x') = x' + \varepsilon a(x'; \varepsilon)$  such that, for all  $|\varepsilon| < \varepsilon_0$ ,*

$$\|\alpha\|_{\xi'} \leq \bar{L} \quad \text{and} \quad \varphi = \text{id} + \varepsilon \alpha : \mathbb{T}_{\xi'}^d \rightarrow \mathbb{T}_{\bar{\xi}}^d. \quad (56)$$

*Furthermore, for any  $(y', x) \in W_{\bar{\xi}}$ ,  $|y' + \varepsilon \beta(y', x)| < \xi$ , so that*

$$\phi_1 = (y' + \varepsilon \beta(y', \varphi(x')), \varphi(x')) : W_{\xi'} \rightarrow W_{\bar{\xi}}, \quad \text{and} \quad \|\phi_1 - \text{id}\|_{\xi'} \leq |\varepsilon| \bar{L}; \quad (57)$$

*finally, the matrix  $1_d + \varepsilon a_x$  is, for any  $x \in \mathbb{T}_{\bar{\xi}}^d$ , invertible with inverse  $1_d + \varepsilon S(x; \varepsilon)$  satisfying*

$$\|S\|_{\bar{\xi}} \leq \frac{\|a_x\|_{\bar{\xi}}}{1 - |\varepsilon| \|a_x\|_{\bar{\xi}}} < \frac{3}{2} \bar{L}, \quad (58)$$

*so that  $\phi_1$  defines a symplectic diffeomorphism.*

The simple proof<sup>18</sup> of this statement is based upon standard tools in mathematical analysis such as the contraction mapping theorem or the inversion

of close-to-identity matrices by Neumann series (see, e.g., Kolmogorov and Fomin (1999)).

From the Lemma and the definition of  $P_1$  in (42), it follows immediately that

$$\|P_1\|_{\xi'} \leq \bar{L}M. \quad (59)$$

Next, by the same technique used to derive (54), one can easily check that

$$\|\tilde{Q}\|_{\xi'}, 2C^2 \left\| \partial_{y'}^2 \tilde{Q}(0, \cdot) \right\|_0 \leq c C^\mu \delta^{-\nu} M = L, \quad (60)$$

for suitable constants  $c \geq \bar{c}$ ,  $\mu \geq \bar{\mu}$ ,  $\nu \geq \bar{\nu}$  (the factor  $2C^2$  has been introduced for later convenience; notice also that  $L \geq \bar{L}$ ). Then, if

$$\varepsilon_0 L := \varepsilon_0 c C^\mu \delta^{-\nu} M \leq \frac{\delta}{3}, \quad (61)$$

there follows that<sup>19</sup>  $\|\tilde{T}\| \leq L$ ; this bound, together with (54), (60), (57), and (59), shows that

$$\begin{cases} |\tilde{E}|, \|\tilde{Q}\|_{\xi'}, \|\tilde{T}\|, \|\phi_1 - \text{id}\|_{\xi'} \leq L \\ \|P_1\|_{\xi'} \leq LM; \end{cases} \quad (62)$$

provided (61) holds (notice that (61) implies (55)).

One step of the iteration has been concluded and the needed estimates obtained. The idea is to iterate the construction infinitely many times, as we proceed to describe.

### Step 3: Iteration and Convergence

In order to iterate Kolmogorov's construction analyzed in Step 2, so as to construct a sequence of symplectic transformations

$$\phi_j : W_{\xi_{j+1}} \rightarrow W_{\xi_j}, \quad (63)$$

closer and closer to the identity, and such that (43) hold, the first thing to do is to choose the sequence  $\xi_j$ : such sequence has to be convergent, so that  $\delta_j = \xi_j - \xi_{j+1}$  has to go to zero rather fast. Inverse power of  $\delta_j$  (which, at the  $j$ th step will play the rôle of  $\delta$  in the previous paragraph) appear in the smallness conditions (see, e.g., (55)): this “divergence” will, however, be beaten by the super-fast decay of  $\varepsilon^{2^j}$ .

Fix  $0 < \xi_* < \xi$  ( $\xi_*$  will be the domain of analyticity of  $\phi_*$  and  $K_*$  in Theorem 1) and, for  $j \geq 0$ , let

$$\begin{cases} \xi_0 := \xi \\ \delta_0 := \frac{\xi - \xi_*}{2} \end{cases} \quad \begin{cases} \delta_j := \frac{\delta_0}{2^j} \\ \xi_{j+1} := \xi_0 - \delta_j = \xi_* + \frac{\delta_0}{2^j} \end{cases} \quad (64)$$

and observe that  $\xi_j \downarrow \xi_*$ . With this choice<sup>20</sup>, Kolmogorov algorithm can be iterated infinitely many times, provided  $\varepsilon_0$  is small enough. To be more precise, let  $c, \mu$  and  $v$  be as in (54), and define

$$C := 2 \max \{ |E|, |\omega|, \|Q\|_\xi, \|T\|, 1 \}. \quad (65)$$

**Smallness assumption:** Assume that  $|\varepsilon| \leq \varepsilon_0$  and that  $\varepsilon_0$  satisfies

$$\begin{aligned} \varepsilon_0 DB \|P\|_\xi \leq 1 \quad \text{where} \quad D := 3c \delta_0^{-(v+1)} C^\mu, \\ B := 2^{v+1}; \end{aligned} \quad (66)$$

notice that the constant  $C$  in (65) satisfies (52) and that (66) implies (55). Then the following claim holds.

**Claim C** Under condition (66) one can iteratively construct a sequence of Kolmogorov symplectic maps  $\phi_j$  as in (63) so that (43) holds in such a way that  $\varepsilon^{2^j} P_j, \Phi_j := \phi_1 \circ \phi_2 \circ \dots \circ \phi_j, E_j, K_j, Q_j$  converge uniformly on  $W_{\xi_*}$  to, respectively,  $0, \phi_*, E_*, K_*, Q_*$ , which are real-analytic on  $W_{\xi_*}$  and  $H \circ \phi_* = K_* = E_* + \omega \cdot y + Q_*$  with  $Q_* = O(|y|^2)$ . Furthermore, the following estimates hold for any  $|\varepsilon| \leq \varepsilon_0$  and for any  $i \geq 0$ :

$$|\varepsilon|^{2^i} M_i := |\varepsilon|^{2^i} \|P_i\|_{\xi_i} \leq \frac{(|\varepsilon| DBM) 2^i}{DB^{i+1}}, \quad (67)$$

$$\begin{aligned} \|\phi_* - \text{id}\|_{\xi_*}, |E - E_*|, \|Q - Q_*\|_{\xi_*}, \|T - T_*\| \\ \leq |\varepsilon| DBM, \end{aligned} \quad (68)$$

Where  $T_* := \left\langle \partial_y^2 Q_*(0, \cdot) \right\rangle^{-1}$ , showing that  $K_*$  is nondegenerate.

**Remark 7** (i) From Claim C Kolmogorov Theorem 1 follows at once. In fact we have proven the following quantitative statement: Let  $\omega \in \mathcal{D}_{\kappa, \tau}^d$  with  $\tau \geq d - 1$  and  $0 < \kappa < 1$ ; let  $Q$  and  $P$  be real-analytic on  $W_\xi = D^d(0, \xi) \times \mathbb{T}_\xi^d$  for some  $0 < \xi \leq 1$  and let  $0 < \theta < 1$ ; let  $T$  and  $C$  be as in, respectively, (51) and (65). There exist  $c_* = c_*(d, \tau, \kappa, \theta) > 1$  and positive integers  $\sigma = \sigma(d, \tau), b$  such that if

$$|\varepsilon| \leq \varepsilon_* := \frac{\xi^\sigma}{c_* \|P\|_\xi C^b} \quad (69)$$

then one can construct a near-to-identity Kolmogorov transformation (Remark 3-(ii))  $\phi_* : W_{\theta\xi} \rightarrow W_\xi$  such that the thesis of Theorem 1 holds together with the estimates

$$\begin{aligned} \|\phi_* - \text{id}\|_{\theta\xi}, |E - E_*|, \|Q - Q_*\|_{\theta\xi}, \|T - T_*\| \\ \leq \frac{|\varepsilon|}{\varepsilon_*} = |\varepsilon| c_* \|P\|_\xi C^b \xi^{-\sigma}. \end{aligned} \quad (70)$$

(The correspondence with the above constants being:  $\xi_* = \theta\xi$ ,  $\delta_0 = \xi(1 - \theta)/2$ ,  $\sigma = v + 1$ ,  $b = \mu + 1$ ,  $D = 3c(2/(1 - \theta))^{v+1} C^{\mu+1}$ ,  $c_* = 3c(4/(1 - \theta))^{v+1}$ ).

(ii) From Cauchy estimates and (68), it follows that  $\|\phi_* - \text{id}\|_{C^p}$  and  $\|Q - Q_*\|_{C^p}$  are small for any  $p$  (small in  $|\varepsilon|$  but not uniformly in<sup>21</sup>  $p$ ).

(iii) All estimates are uniform in  $\varepsilon$  therefore, from Weierstrass theorem (compare note 18), it follows that  $\phi_*$  and  $K_*$  are analytic in  $\varepsilon$  in the complex ball of radius  $\varepsilon_0$ . Analyticity in  $\varepsilon$  and  $\varepsilon$ -power series expansions were very popular in

the XIX and XX century<sup>22</sup>, however, was only J. Moser, within the framework of KAM theory, who proved rigorously (but “indirectly”) for the first time, the convergence of such expansions in 1967: see Moser (1967). Some of this matter is briefly discussed in section “Other Chapters in classical KAM Theory” below.

(iv) **The nearly integrable case.** In Kolmogorov (1954), it is pointed out that Kolmogorov Theorem yields easily the existence of many KAM tori for nearly integrable systems (16) for  $|\varepsilon|$  small enough, provided  $K$  is nondegenerate in the sense that

$$\det K_{yy}(y_0) \neq 0. \quad (71)$$

In fact, without loss of generality, we may assume that  $\omega := H'_0$  is a diffeomorphism on  $B(y_0, 2r)$  and  $\det K_{yy}(y) \neq 0$  for all  $y \in B(y_0, 2r)$ . Furthermore, letting  $B = B(y_0, r)$ , fixing  $\tau > d - 1$  and denoting by  $\ell_d$  the Lebesgue measure on  $\mathbb{R}^d$ , from the remark in note 11 and from the fact that  $\omega$  is a diffeomorphism, there follows that there exists a constant  $c_\#$  depending only on  $d$ ,  $\tau$ , and  $r$  such that

$$\ell_d(\omega(B) \setminus \mathcal{D}_{\kappa, \tau}), \ell_d(\{y \in B : \omega(y) \notin \mathcal{D}_{\kappa, \tau}\}) < c_\# \kappa. \quad (72)$$

Now, let  $B_{\kappa, \tau} := \{y \in B : \omega(y) \in \mathcal{D}_{\kappa, \tau}\}$  (which by (72) has Lebesgue measure  $\ell_d(B_{\kappa, \tau}) \geq \ell_d(B) - c_\# \kappa$ ), then for any  $\bar{y} \in B_{\kappa, \tau}$  we can make the trivial symplectic change of variables  $y \rightarrow \bar{y} + y$ ,  $x \rightarrow x$  so that  $K$  can be written as in (10) with

$$\begin{aligned} E &:= K(\bar{y}), \omega := K_y(\bar{y}), \\ \mathcal{Q}(y, x) &= \mathcal{Q}(y) := K(y) - K(\bar{y}) - K_y(\bar{y}) \cdot y, \end{aligned}$$

(where, for ease of notation, we did not change name to the new symplectic variables) and  $P(\bar{y} + y, x)$  replacing (with a slight abuse of notation)  $P(y, x)$ . By Taylor’s formula,  $\mathcal{Q} = O(|y|^2)$  and, furthermore (since  $\mathcal{Q}(y, x) = \mathcal{Q}(y)$ ,  $\langle \partial_y^2 \mathcal{Q}(0, x) \rangle = \mathcal{Q}_{yy}(0) = K_{yy}(\bar{y})$ , which is invertible according to our hypotheses. Thus,  $K$  is Kolmogorov nondegenerate and Theorem 1 can be

applied yielding, for  $|\varepsilon| < \varepsilon_0$ , a KAM torus  $\mathcal{T}_{\omega, \varepsilon}$ , with  $\omega = K_y(\bar{y})$ , for each  $\bar{y} \in B_{\kappa, \tau}$ . Notice that the measure of initial phase points, which, perturbed, give rise to KAM tori, has a small complementary bounded by  $c_\# \kappa$  (see (72)).

(v) In the nearly integrable setting described in the preceding point, the union of KAM tori is, usually, called the **Kolmogorov set**. It is not difficult to check that the dependence upon  $\bar{y}$  of the Kolmogorov transformation  $\phi_*$  is Lipschitz<sup>23</sup>, implying that the measure of the complementary of Kolmogorov set itself is also bounded by  $\hat{c}_\# \kappa$  with a constant  $\hat{c}_\#$  depending only on  $d$ ,  $\tau$ , and  $r$ .

Indeed, the estimate on the measure of Kolmogorov set can be made more quantitative (i.e., one can see how such estimate depends upon  $\varepsilon$  as  $\varepsilon \rightarrow 0$ ). In fact, revisiting the estimates discussed in **Step 2** above one sees easily that the constant  $c$  defined in (54) has the form<sup>24</sup>

$$c = \hat{c} \kappa^{-4}. \quad (73)$$

where  $\hat{c} = \hat{c}(d, \tau)$  depends only on  $d$  and  $\tau$  (here the Diophantine constant  $\kappa$  is assumed, without loss of generality, to be smaller than one). Thus, the smallness condition (66) reads  $\varepsilon_0 \kappa^{-4} \bar{D} \leq 1$  with some constant  $\bar{D}$  independent of  $\kappa$ : such condition is satisfied by choosing  $\kappa = (\bar{D} \varepsilon_0)^{1/4}$  and since  $\hat{c}_\# \kappa$  was an upper bound on the complementary of Kolmogorov set, we see that *the set of phase points which do not lie on KAM tori may be bounded by a constant times  $\sqrt[4]{\varepsilon_0}$* . Actually, it turns that this bound is not optimal, as we shall see in the next section: see Remark 10.

(vi) The proof of claim C follows easily by induction on the number  $j$  of the iterative steps<sup>25</sup>.

### Arnold’s Scheme

The first detailed proof of Kolmogorov Theorem, in the context of nearly integrable Hamiltonian systems (compare Remark 1-(iii)), was given by V.I. Arnold in 1963.

**Theorem 2 (Arnold 1963a)** *Consider a one-parameter family of nearly integrable Hamiltonians*

$$H(y, x; \varepsilon) := K(y) + \varepsilon P(y, x), \quad (\varepsilon \in \mathbb{R}) \quad (74)$$

with  $K$  and  $P$  real-analytic on  $\mathcal{M} := B(y_0, r) \times \mathbb{T}^d$  (endowed with the standard symplectic form  $dy \wedge dx$ ) satisfying

$$K_y(y_0) = \omega \in \mathcal{D}_{\kappa, \tau}, \quad \det K_{yy}(y_0) \neq 0. \quad (75)$$

Then, if  $\varepsilon$  is small enough, there exists a real-analytic embedding

$$\phi : \theta \in \mathbb{T}^d \rightarrow \mathcal{M} \quad (76)$$

close to the trivial embedding  $(y_0, \text{id})$ , such that the  $d$ -torus

$$\mathcal{T}_{\omega, \varepsilon} := \phi(\mathbb{T}^d) \quad (77)$$

is invariant for  $H$  and

$$\phi_H^t \circ \phi(\theta) = \phi(\theta + \omega t), \quad (78)$$

showing that such a torus is a nondegenerate KAM torus for  $H$ .

**Remark 8** (i) The above Theorem is a corollary of Kolmogorov Theorem 1 as discussed in Remark 7-(iv).

(ii) Arnold's proof of the above Theorem is not based upon Kolmogorov's scheme and is rather different in spirit – although still based on a Newton method – and introduces several interesting technical ideas.

(iii) Indeed, the iteration scheme of Arnold's is more classical and, from the algebraic point of view, easier to construct than Kolmogorov's one, but the estimates involved are somewhat more delicate and introduce a logarithmic correction, so that, in fact, the smallness parameter will be

$$\varepsilon := |\varepsilon| \left( \log |\varepsilon|^{-1} \right)^\rho \quad (79)$$

(for some constant  $\rho = \rho(d, \tau) \geq 1$ ) rather than  $|\varepsilon|$  as in Kolmogorov's scheme, see, also, Remark 9-(iii) and (iv) below.

**Arnold's scheme.** Without loss of generality, one may assume that  $K$  and  $P$  have analytic and bounded extension to  $W_{r, \xi}(y_0) := D(y_0, r) \times \mathbb{T}_\xi^d$  for

some  $\xi > 0$ , where, as above,  $D(y_0, r)$  denotes the complex ball of center  $y_0$  and radius  $r$ . We remark that, in what follows, the analyticity domains of actions and angles play a different rôle.

The Hamiltonian  $H$  in (74) admits, for  $\varepsilon = 0$  the (KAM) invariant torus  $\mathcal{T}_{\omega, 0} = \{y_0\} \times \mathbb{T}^d$  on which the  $K$ -flow is given by  $x \rightarrow x + \omega t$ . Arnold's basic idea is to find a symplectic transformation

$$\begin{aligned} \phi_1 : W_1 &:= D(y_1, r_1) \times \mathbb{T}_{\xi_1}^d \\ &\rightarrow W_0 := D(y_0, r) \times \mathbb{T}_\xi^d, \end{aligned} \quad (80)$$

so that  $W_1 \subseteq W_0$  and

$$\begin{cases} H_1 := H \circ \phi_1 = K_1 + \varepsilon^2 P_1, & K_1 = K_1(y), \\ \partial_y K_1(y_1) = \omega, & \det \partial_y^2 K_1(y_1) \neq 0 \end{cases} \quad (81)$$

(with abuse of notation we denote here the new symplectic variables with the same name of the original variables; as above, dependence on  $\varepsilon$  will, often, not be explicitly indicated). In this way, the initial setup is reconstructed and, for  $\varepsilon$  small enough, one can iterate the scheme so as to build a sequence of symplectic transformations

$$\phi_j : W_j := D(y_j, r_j) \times \mathbb{T}_{\xi_j}^d \rightarrow W_{j-1} \quad (82)$$

so that

$$\begin{cases} H_j := H_{j-1} \circ \phi_j = K_j + \varepsilon^2 P_j, & K_j = K_j(y), \\ \partial_y K_j(y_j) = \omega, & \det \partial_y^2 K_j(y_j) \neq 0. \end{cases} \quad (83)$$

Arnold's transformations, as in Kolmogorov's case, are closer and closer to the identity, and the limit

$$\begin{aligned} \phi(\theta) &:= \lim_{j \rightarrow \infty} \Phi_j(y_i, \theta), \quad \Phi_j := \phi_1 \circ \dots \circ \phi_j : \\ W_j &\rightarrow W_0, \end{aligned} \quad (84)$$

defines a real-analytic embedding of  $\mathbb{T}^d$  into the phase space  $B(y_0, r) \times \mathbb{T}^d$ , which is close to the trivial embedding  $(y_0, \text{id})$ ; furthermore, the torus



$$\mathcal{T}_{\omega,\varepsilon} := \phi(\mathbb{T}^d) = \lim_{j \rightarrow \infty} \Phi_j(y_j, \mathbb{T}^d) \quad (85)$$

is invariant for  $H$  and (78) holds as announced in Theorem 2. Relation (78) follows from the following argument. The radius  $r_j$  will turn out to tend to 0 but in a much slower way than  $\varepsilon^{2j} P_j$ . This fact, together with the rapid convergence of the symplectic transformation  $\Phi_j$  in (84), implies

$$\begin{aligned} \phi_H^t \circ \phi(\theta) &= \lim_{j \rightarrow \infty} \phi_H^t \left( \Phi_j(y_j, \theta) \right) \\ &= \lim_{j \rightarrow \infty} \Phi_j \circ \phi_{H_j}^t(y_j, \theta) \\ &= \lim_{j \rightarrow \infty} \Phi_j(y_j, \theta + \omega t) \\ &= \phi(\theta + \omega t) \end{aligned} \quad (86)$$

(the first equality is just smooth dependence upon initial data of the flow  $\phi_H^t$  together with (84); the second equality is (3); the third equality is due to the fact that  $\phi_{H_j}^t(y_j, \theta) = (y_j, \theta + \omega t) + \epsilon_n$  where  $\epsilon_n$  goes very rapidly to zero and the fourth equality is again (84)).

**Arnold's transformation.** Let us look for a near-to-the-identity transformation  $\phi_1$  so that the first line of (81) holds; such transformation will be determined by a generating function of the form

$$y' \cdot x + \varepsilon g(y', x), \quad \begin{cases} y = y' + \varepsilon g_x(y', x) \\ x' = x + \varepsilon g_{y'}(y', x). \end{cases} \quad (87)$$

Inserting  $y = y' + \varepsilon g_x(y', x)$  into  $H$ , one finds

$$\begin{aligned} H(y' + \varepsilon g_x x) &= K(y') \\ &\quad + \varepsilon [K_y(y') \cdot g_x + P(y', x)] \\ &\quad + \varepsilon^2 (P^{(1)} + P^{(2)}) \end{aligned} \quad (88)$$

with (compare (27))

$$\begin{aligned} P^{(1)} &:= \frac{1}{\varepsilon^2} [K(y' + \varepsilon g_x) - K(y') - \varepsilon K_y(y') \cdot g_x] \\ &= \frac{1}{2} \int_0^1 K_{yy}(y' + t \varepsilon g_x) g_x \cdot \\ g_x dt P^{(2)} &:= \frac{1}{\varepsilon} [P(y' + \varepsilon g_x x) - P(y', x)] \\ &= \int_0^1 P_y(y' + t \varepsilon g_x x) \cdot g_x dt. \end{aligned} \quad (89)$$

**Remark 9** (i) The (naive) idea is to try determine  $g$  so that

$$K_y(y') \cdot g_x + P(y', x) = \text{function of } y' \text{ only}, \quad (90)$$

however, such relation is impossible to achieve. First of all, by taking the  $x$ -average of both sides of (90) one sees that the “function of  $y'$  only” has to be the mean of  $P(y', \cdot)$ , i.e., the zero-Fourier coefficient  $P_0(y')$ , so that the *formal* solution of (90), is (by Fourier expansion)

$$\begin{cases} g = \sum_{n \neq 0} \frac{-P_n(y')}{i K_y(y') \cdot n} e^{in \cdot x}, \\ K_y(y') \cdot g_x + P(y', x) = P_0(y'). \end{cases} \quad (91)$$

But (at difference with Kolmogorov's scheme) the frequency  $K_y(y')$  is a function of the action  $y'$  and since, by the Inverse Function Theorem (Appendix A),  $y \rightarrow K_y(y)$  is a local diffeomorphism, it follows that, in any neighborhood of  $y_0$ , there are points  $y$  such that  $K_y(y) \cdot n = 0$  for some  $n \in \mathbb{Z}^d$ . Thus, in any neighborhood of  $y_0$ , some divisors in (91) will actually vanish and, therefore, *an analytic solution  $g$  cannot exist*<sup>27</sup>.

(ii) On the other hand, since  $K_y(y_0)$  is rationally independent, it is clearly possible (simply by continuity) to control a finite number of divisors in a suitable neighborhood of  $y_0$ , more precisely, for any  $N \in \mathbb{N}$  one can find  $\bar{r} > 0$  such that

$$\begin{aligned} K_y(y) \cdot n &\neq 0, & \forall y \in D(y_0, \bar{r}), \\ &\forall 0 < |n| \leq N; \end{aligned} \quad (92)$$

the important quantitative aspects will be shortly discussed below.

(iii) Relation (90) is also one of the main “identity” in *Averaging Theory* and is related to the so-called *Hamilton–Jacobi equation*. Arnold's proof makes rigorous such theory and shows how a Newton method can be built upon it in order to establish the existence of invariant tori. In a sense, Arnold's approach is much more classical than Kolmogorov's one.

(iv) When (for a given  $y$  and  $n$ ) it occurs that  $K_y(y) \cdot n = K_y(y) \cdot n = 0$ , one speaks of an (exact) *resonance*. As mentioned at the end of point (i), in

the general case, *resonances are dense*. This represents the main problem in Hamiltonian perturbation theory and is a typical feature of *conservative systems*. For generalities on Averaging Theory, Hamilton–Jacobi equation, resonances, etc., see, e.g., Arnold (1974) or Sect. 6.1 and Sect. 6.2 of Arnold et al. (2006).

The key (simple!) idea of Arnold is to split the perturbation in two terms

$$P = \hat{P} + \check{P} \quad \text{where} \quad \begin{cases} \hat{P} := \sum_{|n| \leq N} P_n(y) e^{in \cdot x} \\ \check{P} := \sum_{|n| > N} P_n(y) e^{in \cdot x} \end{cases} \quad (93)$$

choosing  $N$  so that

$$\check{P} = O(\varepsilon) \quad (94)$$

(this is possible because of the fast decay of the Fourier coefficients of  $P$ ; compare (34)). Then, for  $\varepsilon \neq 0$ , (88) can be rewritten as follows

$$\begin{aligned} H(y' + \varepsilon g_x x) &= K(y') \\ &+ \varepsilon [K_y(y') \cdot g_x + \hat{P}(y', x)] \\ &+ \varepsilon^2 (P^{(1)} + P^{(2)} + P^{(3)}) \end{aligned} \quad (95)$$

with  $P^{(1)}$  and  $P^{(2)}$  as in (89) and

$$P^{(3)}(y', x) := \frac{1}{\varepsilon} \check{P}(y', x). \quad (96)$$

Thus, letting<sup>28</sup>

$$g = \sum_{0 < |n| \leq N} \frac{-P_n(y')}{iK_y(y') \cdot n} e^{in \cdot x}, \quad (97)$$

one gets

$$H(y' + \varepsilon g_x x) = K_1(y') + \varepsilon^2 P'(y', x) \quad (98)$$

where

$$\begin{aligned} K_1(y') &:= K(y') + \varepsilon P_0(y'), \\ P'(y', x) &:= P^{(1)} + P^{(2)} + P^{(3)}. \end{aligned} \quad (99)$$

Now, by the IFT (Appendix A), for  $\varepsilon$  small enough, the map  $x \rightarrow x + \varepsilon g_{y'}(y', x)$  can be inverted with a real-analytic map of the form

$$\varphi(y', x'; \varepsilon) := x' + \varepsilon \alpha(y', x'; \varepsilon) \quad (100)$$

so that Arnold's symplectic transformation is given by

$$\phi_1 : (y', x') \rightarrow \begin{cases} y = y' + \varepsilon g_x(y', \varphi(y', x'; \varepsilon)) \\ x = \varphi(y', x'; \varepsilon) = x' + \varepsilon \alpha(y', x'; \varepsilon) \end{cases} \quad (101)$$

(compare (22)). To finish the construction, observe that from the IFT (see Appendix A and the quantitative discussion below) it follows that there exists a (unique) point  $y_1 \in B(y_0, \bar{r})$  so that the second line of (81) holds, provided  $\varepsilon$  is small enough.

In conclusion, the analogous of Proposition 1 holds, describing Arnold's scheme:

**Proposition 2** *If  $\phi_1$  is defined in (101) with  $g$  given in (97) (with  $N$  so that (94) holds) and  $\varphi$  given in (100), then (81) holds with  $K_1$  as in (99) and  $P_1(y', x') := P'(y', \varphi(y', x'))$  with  $P'$  defined in (99), (96), and (89).*

**Estimates and convergence.** If  $f$  is a real-analytic function with analytic extension to  $W_{r, \xi}$ , we denote, for any  $r' \leq r$  and  $\xi' \leq \xi$ ,

$$\|f\|_{r', \xi'} := \sup_{W_{r', \xi'}(y_0)} |f(y, x)|; \quad (102)$$

furthermore, we define

$$T := K_{yy}(y_0)^{-1}, \quad M := \|P\|_{r, \xi}, \quad (103)$$

and assume (without loss of generality)

$$\begin{aligned} \kappa < 1, \quad r < 1, \quad \xi \leq 1, \\ \max \left\{ 1, \|K_y\|_r, \|K_{yy}\|_{r'}, \|T\| \right\} < C, \end{aligned} \quad (104)$$

for a suitable constant  $C$  (which, as above, will not change during the iteration).

We begin by discussing how  $N$  and  $\bar{r}$  depend upon  $\varepsilon$ . From the exponential decay of the Fourier coefficients (34), it follows that, choosing

$$N := 5\delta^{-1}\lambda, \quad \text{where} \quad \lambda := \log|\varepsilon|^{-1}, \quad (105)$$

then

$$\|\check{P}\|_{r, \xi - \frac{\delta}{2}} \leq |\varepsilon|M \quad (106)$$

provided

$$|\varepsilon| \leq \text{const} \delta^{4d} \quad (107)$$

for a suitable<sup>29</sup>  $\text{const} = \text{const}(d)$ .

The second key inequality concerns the control of the small divisors  $K_y(y') \cdot n$  appearing in the definition of  $g$  (see (97)), in a neighborhood  $D(y_0, \bar{r})$  of  $y_0$ : this will determine the size of  $\bar{r}$ .

Recalling that  $K_y(y_0) = \omega \in \mathcal{D}_{\tau, \kappa}$ , by Taylor's formula and (9), one finds, for any  $0 < |n| \leq N$  and any  $y' \in D(y_0, \bar{r})$ ,

$$\begin{aligned} |K_y(y') \cdot n| &= |\omega \cdot n + (K_y(y') - K_y(y_0)) \cdot n| \\ &\geq |\omega \cdot n| \left( 1 - \frac{\|K_{yy}\|_r}{|\omega \cdot n|} |n| \bar{r} \right) \\ &\geq \frac{\kappa}{|n|^\tau} \left( 1 - \frac{C}{\kappa} |n|^{\tau+1} \bar{r} \right) \\ &\geq \frac{\kappa}{|n|^\tau} \left( 1 - \frac{C}{\kappa} N^{\tau+1} \bar{r} \right) \\ &\geq \frac{1}{2} \frac{\kappa}{|n|^\tau}, \end{aligned} \quad (108)$$

provided  $\bar{r} \leq r$  satisfies also

$$\bar{r} \leq \frac{\kappa}{2CN^{\tau+1}} \stackrel{(105)}{=} \frac{\kappa}{2 \cdot 5^{\tau+1} C (\delta^{-1}\lambda)^{\tau+1}}. \quad (109)$$

Equation (108) allows easily to control Arnold's generating function  $g$ . For example:

$$\begin{aligned} \|g_x\|_{\bar{r}, \xi - \frac{\delta}{2}} &= \sup_{D(y_0, \bar{r}) \times \mathbb{T}^{\kappa} \times \mathbb{T}^{\kappa}} \left| \sum_{0 < |n| \leq N} \frac{n P_n(y')}{K_y(y') \cdot n} e^{in \cdot x} \right| \\ &\times \leq \sum_{0 < |n| \leq N} \frac{\sup_{D(y_0, r)} |P_n(y')|}{|K_y(y') \cdot n|} |n| e^{(\xi - \frac{\delta}{2}) |n|} \\ &\leq \sum_{n \in \mathbb{Z}^d} M \frac{2|n|^\tau}{\kappa} e^{-\frac{\delta}{2} |n|} \\ &\leq \text{const} \frac{M}{\kappa} \delta^{-(\tau+1+d)}, \end{aligned} \quad (110)$$

where “const” denotes a constant depending on  $d$  and  $\tau$  only; compare also Remark 6-(i). Let us now discuss, from a quantitative point of view, how to choose the new “center” of the action variables  $y_1$ , which is determined by the requirements in (81). Assuming that

$$\bar{r} \leq \frac{r}{2} \quad (111)$$

(allowing to use Cauchy estimates for  $y$ -derivatives of  $K$  or  $P$  in  $D(y_0, \bar{r})$ ), it is not difficult to see that the quantitative IFT of Appendix A implies that there exists a unique  $y_1 \in D(y_0, \bar{r})$  such that (81) holds. In fact, assuming

$$\begin{cases} 8C^2 \frac{\bar{r}}{r} \leq 1, \\ \frac{8CM}{r\bar{r}} |\varepsilon| \leq 1 \end{cases} \quad (112)$$

one can show that<sup>30</sup>

$$|y_1 - y_0| \leq 4CMr^{-1} |\varepsilon| \leq \frac{\bar{r}}{2}, \quad (113)$$

and

$$\begin{aligned} \partial_y^2 K_1(y_1) &:= K_{yy}(y_1) \\ &+ \varepsilon \partial_y^2 P_0(y_1) =: T^{-1}(1_d + A) \end{aligned} \quad (114)$$

with  $A := T(K_{yy}(y_1) - K_{yy}(y_0) + \varepsilon \partial_y^2 P_0(y_1))$  satisfying

$$\|A\| \leq \frac{12C^3M}{r^2} |\varepsilon| \stackrel{(112)}{\leq} \frac{12C^3M}{r} \frac{1}{8C^2\bar{r}} |\varepsilon| \stackrel{(112)}{\leq} \frac{3}{16}. \quad (115)$$

Equations (114) and (115) imply that  $\partial_y^2 K(y_1)$  is invertible (Neumann series) and that<sup>31</sup>

$$\partial_y^2 K_1(y_1)^{-1} = T + \varepsilon \tilde{T}, \quad \|\tilde{T}\| \leq 15 \frac{C^4 M}{r^2}. \quad (116)$$

Finally, notice that by (113),

$$\frac{\|g_x\|_{\bar{r}, \bar{\xi}}}{r}, \|g_{y'}\|_{\bar{r}, \bar{\xi}}, \frac{\|\tilde{y}\|}{r}, \|\tilde{K}_y\|_{\bar{r}}, \|\tilde{K}_{yy}\|, \|\tilde{T}\| \leq c\kappa^{-2} C^\mu \delta^{-\nu} \lambda^\rho M^{\varepsilon=L}, \|P'\|_{\bar{\xi}} \leq c\kappa^{-2} C^\mu \delta^{-\nu} \lambda^\rho M^2 = LM, \quad (119)$$

where  $c = c(d, \tau) > 1$ ,  $\mu \in \mathbb{Z}_+$ ,  $\nu$  and  $\rho$  are positive integers depending on  $d$  and  $\tau$ . Now, by<sup>32</sup> Lemma 1 and (119), one has that map  $x \rightarrow x + \varepsilon g_y(y', x)$  has, for any  $y' \in D_{\bar{r}}(y_0)$ , an analytic inverse  $\varphi = x' + \varepsilon \mathcal{X}(x'; y', \varepsilon) =: \varphi(y', x')$  on  $\mathbb{T}_{\bar{\xi}-\frac{\delta}{3}}^d$  provided (55) holds (with  $L$  as in (119)), in which case (56) holds (for any  $|\varepsilon| \leq \varepsilon_0$  and any  $y' \in D_{\bar{r}}(y_0)$ ). Furthermore, under the above hypothesis, it follows that<sup>33</sup>

$$\left\{ \begin{array}{l} \phi_1 := (y' + \varepsilon g_x(y', \varphi(y', x')), \varphi(y', x')) : W_{\bar{r}/2, \bar{\xi}-\delta}(y_1) \rightarrow W_{r, \xi}(y_0) \\ \|\phi_1 - \text{id}\|_{\bar{r}/2, \bar{\xi}-\delta} \leq |\varepsilon| L. \end{array} \right. \quad (120)$$

Finally, letting  $P_1(y', x') := P'(y', \varphi(y', x'))$  one sees that  $P_1$  is real-analytic on  $W_{\bar{r}/2, \bar{\xi}-\delta}(y_1)$  and bounded on such domain by

$$\|P_1\|_{\bar{r}/2, \bar{\xi}-\delta} \leq LM. \quad (121)$$

In order to iterate the above construction, we fix  $0 < \zeta_* < \zeta$  and set

$$C := 2 \max \left\{ 1, \|K_y\|_r, \|K_{yy}\|_r, \|T\| \right\}, \quad \gamma := 3C, \\ \delta_0 := \frac{(\gamma - 1)(\zeta - \zeta_*)}{\gamma}; \quad (122)$$

$$D(y_1, \bar{r}/2) \subseteq D(y_0, \bar{r}). \quad (117)$$

Now, all the estimating tools are set up and, writing

$$K_1 := K + \varepsilon \tilde{K} = K + \varepsilon P_0(y'), y_1 := y_0 + \varepsilon \tilde{y}, \quad (118)$$

one can easily prove (along the lines that led to (54)) the following estimates, where as in section “Kolmogorov Theorem,”  $\bar{\xi} := \xi - \frac{2}{3}\delta$  and  $\bar{r}$  is as above:

$\xi_j$  and  $\delta_j$  as in (64) but with  $\delta_0$  as in (122); we also define, for any  $j \geq 0$ ,

$$\lambda_j := 2^j \lambda = \log \varepsilon_0^{-2^j}, \quad r_j := \frac{\kappa}{4 \cdot 5^{\tau+1} C \left( \delta_j^{-1} \lambda_j \right)^{\tau+1}}; \quad (123)$$

(this part is adapted from **Step 3** in section “Kolmogorov Theorem”; see, in particular, (104)). With such choices, it is not difficult to check that the iterative construction may be carried out infinitely many times yielding, as a byproduct, Theorem 2 with  $\phi$  real-analytic on  $\mathbb{T}_{\xi_*}^d$ , provided  $|\varepsilon| \leq \varepsilon_0$  with  $\varepsilon_0$  satisfying<sup>34</sup>

$$\left\{ \begin{array}{ll} \varepsilon_0 \leq e^{-\beta} & \text{with } \beta := \frac{\delta_0}{5} \left( \frac{\kappa}{Cr} \right)^{\frac{1}{\tau+1}} \\ \varepsilon_0 DB \|P\|_{r, \xi} \leq 1 & \text{with } D := 3c\kappa^{-2} \delta_0^{-(\nu+1)} C^{\mu+1}, \\ B := \gamma^{\nu+1} (\log \varepsilon_0^{-1})^\rho. \end{array} \right. \quad (124)$$

**Remark 10** Notice that the power of  $\kappa^{-1}$  (the inverse of the Diophantine constant) in the second smallness condition in (124) is two, which implies (compare Remark 7-(v)) that the measure of the complementary of Kolmogorov set may be bounded by a constant times  $\sqrt{\varepsilon}$ , where  $\varepsilon :=$

$\varepsilon(\log \varepsilon)^{-\rho}$ . This bound is almost optimal (i.e., optimal, up to logarithmic corrections) as the trivial example  $(y_1^2 + y_2^2)/2 + \varepsilon \cos(x_1)$  shows: such Hamiltonian is integrable and the phase portrait shows that the separatrices of the pendulum  $y_1^2/2 + \varepsilon \cos x_1$  bound a region of area  $\sqrt{|\varepsilon|}$  with no KAM tori (as the librational curves within such region are not graphs over the angles).

Taking out the logarithm is not a completely trivial matter, and even though in the literature is normally claimed that the sharp estimate holds, a complete proof of this fact is hard to find. For a recent detailed proof, see Biasco and Chierchia (2018).

### The Differentiable Case: Moser's Theorem

J.K. Moser, in 1962, proved a perturbation (KAM) Theorem, in the framework of area-preserving twist mappings of an annulus<sup>35</sup>  $[0, 1] \times \mathbb{S}^1$ , for integrable analytic systems perturbed by a  $C^k$  perturbation (Moser 1961, 1962). Moser's original set up corresponds to the Hamiltonian case with  $d = 2$  and the required smoothness was  $C^k$  with  $k = 333$ . Later, this number was brought down to 5 by H. Rüssmann (Rüssmann 1970).

Moser's original approach, similarly to the approach that led J. Nash to prove its theorem on the smooth embedding problem of compact Riemannian manifolds, Nash (1956), is based on a smoothing technique (via convolutions), which re-introduces at each step of the Newton iteration a certain number of derivatives which one loses in the inversion of the small divisor operator.

The technique, which we shall describe here, is again due to Moser (1970) but is rather different from the original one, and it is based on a quantitative analytic KAM Theorem (in the style of statement in Remark 7-(i) above) in conjunction with a characterization of differentiable functions in terms of functions, which are real-analytic on smaller and smaller complex strips; see Moser (1966) and, for an abstract functional approach, Zehnder (1975, 1976). By the way, this approach, suitably refined, leads to optimal differentiability assumptions (i.e., the Hamiltonian may be assumed to be  $C^\ell$  with  $\ell > 2d$ ); see, Pöschel (1982) and the beautiful

exposition Salamon (2004), which inspires the presentation reported here.

Let us consider a Hamiltonian  $H = K + \varepsilon P$  (as in (18)) with  $K$  a real-analytic Kolmogorov normal form as in (10) with  $\omega \in \mathcal{D}_{\kappa, \tau}$  and  $Q$  real-analytic;  $P$  is assumed to be a  $C^\ell(\mathbb{R}^d \times \mathbb{T}^d)$  function with  $\ell = \ell(d, \tau)$  to be specified later<sup>36</sup>.

**Remark 11** The analytic KAM theorem, we shall refer to, is the quantitative Kolmogorov Theorem as stated in Remark 7-(i) above, with (70) strengthened by including in the left hand side of (70) also<sup>37</sup>  $\|\partial(\phi_* - \text{id})\|_{\theta\xi}$  and  $\|\partial(Q - Q_*)\|_{\theta\xi}$  (where “ $\partial$ ” denotes, here, “Jacobian” with respect to  $(y, x)$  for  $(\phi_* - \text{id})$  and “gradient” for  $(Q - Q_*)$ ).

The analytic characterization of differentiable functions, suitable for our purposes, is explained in the following two lemmata<sup>38</sup>.

**Lemma 2 (Jackson, Moser, Zehnder)** *Let  $f \in C^l(\mathbb{R}^d)$  with  $l > 0$ . Then, for any  $0 < \xi \leq 1$  there exists a real-analytic function  $f_\xi : X_\xi^d := \{x \in \mathbb{C}^d : |\text{Im}x_j| < \xi\} \rightarrow \mathbb{C}$  such that*

$$\begin{cases} \sup_{X_\xi^d} |f_\xi| \leq c \|f\|_{C^0}, & \|f_\xi - f\|_{C^s} \leq c \|f\|_{C^l} \xi^{l-s}, (s \in \mathbb{N}, s \leq l) \\ \sup_{X_{\xi'}^d} |f_\xi - f_{\xi'}| \leq c \|f\|_{C^l} \xi^l, & \forall 0 < \xi' < \xi, \end{cases} \quad (125)$$

where  $c = c(d, l)$  is suitable constant; if  $f$  is periodic in some variable  $x_p$ , so is  $f_\xi$ .

**Lemma 3 (Bernstein, Moser)** *Let  $l \in \mathbb{R}_+ \setminus \mathbb{Z}$ ; let  $f_0 = 0$  and let, for any  $j \geq 1$ ,  $f_j$  be real analytic functions on  $X_j^d := \{x \in \mathbb{C}^d : |\text{Im}x_k| < 2^{-j}\}$  such that*

$$\sup_{X_j^d} |f_j - f_{j-1}| \leq A 2^{-jl} \quad (126)$$

for some constant  $A$ . Then,  $f_j$  tends uniformly on  $\mathbb{R}^d$  to a function  $f \in C^l(\mathbb{R}^d)$  such that, for a suitable constant  $C = C(d, l) > 0$ ,

$$\|f\|_{C^l(\mathbb{R}^d)} \leq CA. \quad (127)$$

Finally, if the  $f_i$ 's are periodic in some variable  $x_k$  then so is  $f$ .

Now, denote by  $X_\xi = X_\xi^d \times \mathbb{T}^d \subset \mathbb{C}^{2d}$  and define (compare Lemma 2)

$$P^j := P_{\xi_j}, \quad \xi_j := \frac{1}{2^j}. \quad (128)$$

**Claim M** *If  $|\varepsilon|$  is small enough and if  $\ell > \sigma + 1$ , then there exists a sequence of Kolmogorov symplectic transformations  $\{\Phi_j\}_{j \geq 0}$ ,  $|\varepsilon|$ -close to the identity, and a sequence of Kolmogorov normal forms  $K_j$  such that*

$$H_j \circ \Phi_j = K_{j+1} \quad \text{on} \quad W_{\xi_{j+1}} \quad (129)$$

where

$$\begin{aligned} H_j &:= K + \varepsilon P^j \\ \Phi_0 &= \phi_0 \quad \text{and} \quad \Phi_j := \Phi_{j-1} \circ \phi_j, (j \geq 1) \\ \phi_j : W_{\xi_{j+1}} &\rightarrow W_{\alpha \xi_j}, \quad \Phi_{j-1} : W_{\alpha \xi_j} \rightarrow X_{\xi_j}, \quad j \geq 1 \\ \text{and} \quad \alpha &:= \frac{1}{\sqrt{2}}, \quad \sup_{x \in \mathbb{T}_{\xi_{j+1}}^d} |\Phi_j(0, x) - \Phi_{j-1}(0, x)| \\ &\leq \text{const} |\varepsilon| 2^{-(\ell-\sigma)j}. \end{aligned} \quad (130)$$

The proof of Claim M follows easily by induction<sup>39</sup> from Kolmogorov Theorem (compare Remark 11) and Lemma 2.

From Claim M and Lemma 3 (applied to  $f_j(x) = \Phi_j(0, x) - \Phi_0(0, x)$  and  $l = \ell - \sigma$ , which may be assumed not integer), it then follows that  $\Phi_j(0, x)$  converges in the  $C^1$  norm to a  $C^1$  function  $\phi: \mathbb{T}^d \rightarrow \mathbb{R}^d \times \mathbb{T}^d$ , which is  $\varepsilon$ -close to the identity, and, because of (129),

$$\begin{aligned} \phi(x + \omega t) &= \lim \Phi_j(0, x + \omega t) \\ &= \lim \phi_{H_j}^t \circ \Phi_j(0, x) \\ &= \phi_H^t \circ \phi(x) \end{aligned} \quad (131)$$

showing that  $\phi(\mathbb{T}^d)$  is a  $C^1$  KAM torus for  $H$  (note that the map  $\phi$  is close to the trivial embedding  $x \rightarrow (0, x)$ ).

### Lower Dimensional KAM Tori

We consider the existence of quasi-periodic solutions with a number of frequencies smaller than the number of degrees of freedom<sup>49</sup>. Such solutions span *lower dimensional* (non-Lagrangian) *tori*.

Certainly, this is one of the most important topics in modern KAM theory, not only in view of applications to classical problems, but especially in view of extensions to infinite dimensional systems, namely, PDEs (Partial Differential Equations) with a Hamiltonian structure. For a review on lower dimensional tori (in finite dimensions), we refer the reader to Sevryuk (2003).

In 1965 V.K. Melnikov (1965) stated a precise result concerning the persistence of *stable* (or “elliptic”) lower dimensional tori; the hypotheses of such result are, now, commonly referred to as “Melnikov conditions.” However, a proof of Melnikov’s statement was given only later by Moser (1967) for the case  $n = d - 1$  and, in the general case, by H. Eliasson in (1988) and, independently, by S.B. Kuksin (1988). The *unstable* (“partially hyperbolic”) case (i.e., the case for which the lower dimensional tori are linearly unstable and lie in the intersection of stable and unstable Lagrangian manifolds) is simpler and a complete perturbation theory was already given in Graff (1974), Moser (1967) and Zehnder (1976) (roughly speaking, the normal frequencies to the torus do not resonate with the inner (or “proper”) frequencies associated to the quasi-periodic motion). Since then, Melnikov conditions have been significantly weakened and a lot of technical progress has been done; see Sevryuk (2003), Sects. 5, 6, and 7, and references therein.

As an example we consider a system with  $n + m$  degrees of freedom with Hamiltonian

$$H = K(x, y, z; \xi) + \varepsilon P(x, y, z; \xi) \quad (132)$$

where  $(x, y) \in \mathbb{T}^n \times \mathbb{R}^n$ ,  $z = \frac{1}{\sqrt{2}}(p + iq) \in \mathbb{C}^m$  are pairs of standard symplectic coordinates, while  $\xi$  is a real parameter running over a compact set  $\Pi \subset \mathbb{R}^n$  of positive Lebesgue measure<sup>50</sup>.  $K, P$  are Lipschitz in  $\xi$  and analytic with respect to the dynamical variables  $x, y, z, \bar{z}$  (note that when we complexify  $(p, q)$  the variables  $z, \bar{z}$  become independent) in the complexified domain

$$(x, y, z, \bar{z}) \in D(s, r) := \mathbb{T}_s^n \times B_{r,2}(\mathbb{C}^n) \times B_r(\mathbb{C}^m) \times B_r(\mathbb{C}^m)$$

namely, they can be written in totally convergent Taylor Fourier series as

$$K = \sum_{d \in \mathbb{N}} K^{(d)} = \sum_{d \in \mathbb{N}} \sum_{\substack{\ell \in \mathbb{Z}^n, l \in \mathbb{N}^n, \alpha, \beta \in \mathbb{N}^m \\ 2|\ell| + |\alpha| + |\beta| = d}} K_{\ell, \alpha, \beta} e^{i\ell \cdot x} y^\alpha z^\beta, \quad K = K^{(0)}(\xi) + \omega^{(0)}(\xi) \cdot y + \sum_{j=1}^m \Omega_j^{(0)}(\xi) |z_j|^2 + K^{(\geq 3)} \quad (133)$$

where  $y^l z^\alpha \bar{z}^\beta = \prod_{i=1}^n y_i^{l_i} \prod_{j=1}^m z_j^{\alpha_j} \bar{z}_j^{\beta_j}$ .

Here the apex  $(d)$  denotes the homogeneous components of *degree*  $d$ , provided that we assign degree two to the variables  $y$ , degree one to the variables  $z$ , and degree zero to the variables  $x$ . Note that this choice of degrees is the one that makes the symplectic form homogeneous of degree two, since the variables  $x$ , which are not close to zero, must have degree zero.

We shall assume that, for all  $\xi \in \Pi$ ,  $K$  admits the  $n$ -torus

$$\mathcal{T}_0(\xi) := \{y = 0\} \times \mathbb{T}^n \times \{z = 0\}$$

as a linearly stable invariant torus and is written in normal form

here  $K^{(\geq 3)}$  is an analytic Hamiltonian with minimal degree at least three while  $K^{(0)}(\xi)$  is a constant. The  $\phi_K^t$  flow decouples in the linear flow  $x \in \mathbb{T}^n \rightarrow x + \omega^{(0)}(\xi)t$  times the motion of  $m$  (decoupled) harmonic oscillators with characteristic frequencies  $\Omega_j^{(0)}(\xi)$  (sometimes referred to as *normal frequencies*).

We have the following result:

**Theorem 3 (Pöschel 1989)** *Fix  $\gamma > 0, \tau > n$  then for all  $|\varepsilon|$  sufficiently small there exists Lipschitz functions  $\omega(\xi), \Omega(\xi): \Pi \rightarrow \mathbb{R}^{n+m}$   $\varepsilon$ -close to  $\omega^{(0)}(\xi), \Omega^{(0)}(\xi)$  such that setting*

$$\Pi^* := \left\{ \xi \in \Pi : |\omega \cdot \ell + k \cdot \Omega| \geq \frac{\gamma}{|\ell|^\tau}, \forall (\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\} : |k| \leq 2 \right\} \quad (134)$$

then for all  $\xi \in \Pi^*$  there exists a change of variables  $\Phi$ ,  $\varepsilon$ -close to the identity, such that

$$H \circ \Phi = \omega(\xi) \cdot y + \sum_{j=1}^m \Omega_j(\xi) |z_j|^2 + H^{(\geq 3)} \quad (135)$$

namely, it is in normal form with frequencies  $\omega(\xi), \Omega(\xi)$ .

Now in order to make this result interesting, we have to give conditions which ensure that the set  $\Pi^*$  has positive Lebesgue measure. This follows, for instance, by requiring that  $\xi \rightarrow \omega^{(0)}(\xi)$  is a Lipeomorphism and that the Melnikov conditions hold. Explicitly, for any  $(\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\}$  with  $|k| \leq 2$ , we define:

$$\mathcal{R}_{\ell, k}^{(0)} := \{\xi \in \Pi : \omega^{(0)} \cdot \ell + k \cdot \Omega^{(0)} = 0\}, \quad \text{and assume that } |\mathcal{R}_{\ell, k}^{(0)}| = 0, \mathcal{R}_{0, k}^{(0)} = \emptyset \quad (136)$$

This formulation has been borrowed from Pöschel (1989), to which we refer for a complete proof; the description of the set  $\Pi^*$  in terms of the

*final frequencies* is the one given in Berti and Biasco (2011); for the differentiable analog, see Chierchia and Qian (2004).



In order to give a sketch of the proof, let us introduce some notation: we define the degree projections  $\Pi^{\leq j}, \Pi^j, \Pi^{> j}$  as

$$\Pi^j H = H^{(j)}, \quad \Pi^{\leq j} H = \sum_{0 \leq d \leq j} H^{(d)},$$

in the same way  $H^{\geq 3}$  is a Hamiltonian with minimal degree at least three, while  $H^{\leq 2}$  is a polynomial Hamiltonian of maximal degree two, etc.

We endow the space of Hamiltonians with a structure of scale of Banach spaces with respect to the norm defined as follows.

We represent a vector field on  $\mathbb{R}^{2n} \times \mathbb{C}^m$  as

$$X = \sum_{i=1}^n X^{(x_i)}(x, y, z) \frac{\partial}{\partial x_i} + X^{(y_i)}(x, y, z) \frac{\partial}{\partial y_i} + \sum_{j=1}^m X^{(z_j)}(x, y, z) \frac{\partial}{\partial z_j},$$

where each component is an analytic function,

$$X^{(v)}(x, y, z) = \sum_{\ell \in \mathbb{Z}^n, j \in \mathbb{N}^n, \alpha, \beta \in \mathbb{N}^m} X_{\ell, \alpha, \beta}^{(v)} e^{i\ell \cdot x} y^j z^\alpha \bar{z}^\beta, \\ v = x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_m.$$

Finally we define the majorant vector field  $\underline{X}$  by setting  $\underline{X}_{\ell, \alpha, \beta}^{(v)} = |X_{\ell, \alpha, \beta}^{(v)}|$ .

$$|H|_{s, r} := \sup_{D(s, r)} \left( |H| + |\underline{X}_H^{(x)}|^{\text{lip}} + \frac{1}{r^2} |\underline{X}_H^{(z)}|^{\text{lip}} + \frac{1}{r} |\underline{X}_H^{(z)}|^{\text{lip}} \right) \quad (137)$$

where given a Lipschitz map  $f: \Pi \rightarrow E$  with  $E$  a Banach space, we denote by  $|f|_E^{\text{lip}}$  the *inhomogeneous Lipschitz norm*

$$|f|_E^{\text{lip}} := \sup_{\xi \in \Pi} |f(\xi)|_E + \sup_{\xi \neq \xi' \in \Pi} \frac{|f(\xi) - f(\xi')|_E}{|\xi - \xi'|}$$

This norm is less natural than the one defined in (46), in particular due to the presence of the majorant it is not coordinate independent.

However, it is closed with respect to Poisson brackets, projection onto the components of homogeneous degree, and has exponentially small smoothing estimates for the ultraviolet terms, i.e., has properties similar to **(P1)–(P5)**; moreover, it turns out that with this definition the smallness assumptions on  $\varepsilon$  in the KAM theorem 3 are independent of  $m$ . Now our goal is to find:

- (A) A sequence  $\omega^{(n)}, \Omega^{(n)}$  defined and Lipschitz for  $\xi \in \Pi$  and tending to  $\omega, \Omega$  super-exponentially
- (B) A sequence  $\varepsilon_n$  rapidly converging to zero and a (rapidly converging) sequence of changes of variables  $\Psi_n$ , well defined and Lipschitz for  $\xi$  in a nested sequence of domains which contains  $\Pi^*$  defined by

$$\Pi_n^* := \left\{ \xi \in \Pi_{n-1}^* : \left| \omega^{(n)} \cdot \ell + k \cdot \Omega^{(n)} \right| \geq \frac{\gamma}{|\ell|^\tau}, \forall (\ell, k) \in \mathbb{Z}^{n+m} \setminus \{0\} : \right. \\ \left. |k| \leq 2, \quad |\ell| \leq N_n \right\}, \quad (138)$$

with  $N_n \sim \ln(\varepsilon_n^{-1})$ , such that

$$H_0 = K_0 + \varepsilon_0 P_0,$$

$$H_{n+1} = \Psi_n H_n = K_{n+1} + \varepsilon_{n+1} P_{n+1},$$

$K_n = K_n^{(0)} + K_n^{(2)} + K_n^{(\geq 3)}$  with  $K_n^{(0)}$  depending only on  $\xi$ ,

$$K_n^{(2)} = \omega^{(n)} \cdot y + \sum_j \Omega_j^{(n)} |z_j|^2,$$

Finally  $P_n$  is a polynomial of maximal degree two and  $\varepsilon_n \sim \varepsilon_{n-1}^{3/2}$  tends to zero super-exponentially.

Let us show how to perform one step of this procedure. We claim that  $\Psi_n$  is the time-one flow of a generating function  $S_n \sim O(N_n^{3\tau} \varepsilon_n)$ , where the closeness in the norm (137) with an

appropriate choice of parameters  $s_n, r_n$ . Recalling the Lie exponentiation formula, we have

$$\begin{aligned} H_n \circ \Psi_n &= H_n + \{S_n, H_n\} + O(\varepsilon_n^2) \\ &= K_n + \varepsilon_n P_n + \{S_n, K_n\} + O(N_n^{6\tau} \varepsilon_n^2). \end{aligned}$$

Our goal is achieved provided that we fix  $S_n$  so that

$$\begin{aligned} \Pi^{(\leq 2)}(K_n + \varepsilon_n P_n + \{S_n, K_n\}) \\ = \Pi^{(\leq 2)}K_{n+1} + O(\varepsilon_n^{3/2}), \end{aligned}$$

recall that we only want the terms of degree at most two to be in normal form, this is why we apply the projection  $\Pi^{(\leq 2)}$ . We assume that  $S_n$  is a polynomial of maximal degree two and solve the equations above in increasing homogeneous degrees, recalling that  $\{F^{(d_1)}, G^{(d_2)}\}$  has degree  $d_1 + d_2 - 2$ . We get a triangular system:

$$\begin{aligned} \{S_n^{(0)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(0)} + K_{n+1}^{(0)} - K_n^{(0)} + O(\varepsilon_n^{3/2}) \\ \{S_n^{(1)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(1)} - \{S_n^{(0)}, K_n^{(3)}\} + O(\varepsilon_n^{3/2}) \\ \{S_n^{(2)}, K_n^{(2)}\} &= -\varepsilon_n P_n^{(2)} - \{S_n^{(0)}, K_n^{(4)}\} - \{S_n^{(1)}, K_n^{(3)}\} \\ &\quad + K_{n+1}^{(2)} - K_n^{(2)} + O(\varepsilon_n^{3/2}). \end{aligned} \quad (139)$$

which we solve for  $\xi \in \Pi_n$ , just like we did for Eq. (31), by noticing that

$$\begin{aligned} \{K_n^{(2)}, e^{i\ell \cdot x} y^j z^\alpha \bar{z}^\beta\} \\ = i(\omega^{(n)} \cdot \ell + \Omega^{(n)} \cdot (\alpha - \beta)) e^{i\ell \cdot x} y^j z^\alpha \bar{z}^\beta \end{aligned}$$

and that all the ultraviolet terms with frequency  $\geq N_n$  can be ignored since they are  $O(\varepsilon_n^{3/2})$ .

One can also consider more general cases, for instance, where the conditions (136) hold only for  $(\ell, k)$  with  $|k| \leq 1$ , namely, the *second* Melnikov conditions do not hold. Then one can still prove the existence of a torus, for  $\xi$  in some positive measure Cantor-like set, this was done by J. Bourgain in 1997. We state his theorem (written with our notations):

**Theorem 4 (Bourgain 1997)** *Let  $H(x, y, z)$  be of the form*

$$H = \omega^{(0)}(\xi) \cdot y + \frac{1}{2}|y|^2 + \sum_{j=1}^m \Omega_j^{(0)}(\xi) |z_j|^2 + \varepsilon P$$

where we assume that  $\omega^{(0)}$  is diophantine and that condition (136) holds for  $(\ell, k)$  with  $|k| \leq 1$ . Then, for any fixed small  $\varepsilon > 0$  and for  $\lambda$  taken in a set of positive measure, there exists a perturbed torus with frequency vector  $\omega = \lambda \omega^{(0)}$ , parametrized as

$$x = \omega t + X(\omega t), \quad y = Y(\omega t), \quad z = Z(\omega t)$$

with  $(X, Y, Z)$  quasi-periodic and of size, say,  $O(\varepsilon^{1/2})$  in a suitable real analytic function space norm.

We remark that in general in this case one does not have information on the stability in the  $z$  directions.

Bourgain's approach to this problem was to look directly for the quasi-periodic solution. This amounts to looking for a map

$$\begin{aligned} \mathbf{i} : \mathbb{T}^n &\rightarrow \mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^m, \\ \varphi &\rightarrow \mathbf{i}(\varphi) = (\varphi + X(\varphi), Y(\varphi), Z(\varphi)), \end{aligned}$$

and for a frequency  $\omega \in \mathbb{R}^n$ , which solve the functional equation

$$\mathcal{F}(\mathbf{i}) := \omega \cdot \partial_\varphi \mathbf{i}(\varphi) - X_H(\mathbf{i}(\varphi)) = 0.$$

Now in order to solve this functional problem, we apply a Nash-Moser quadratic algorithm, starting from the approximate solution  $\mathbf{i} = \mathbf{i}_0(\varphi) = (\varphi, 0, 0)$  and  $\omega = \omega^{(0)}$  and constructing a super-exponentially convergent sequence of approximate solutions  $\mathbf{i}_n(\varphi), \omega^{(n)}$ . The key point is to invert (with some quantitative control on the bounds) the linearized operator at an approximate solution. This is in general a much more difficult task with respect to solving the homological Eqs. (139), since it involves a linear operator which depends quasi-periodically on time. In the case of maximal tori, this problem can be overcome, e.g., Celletti and Chierchia (1988), by exploiting the symplectic structure. In the more

difficult elliptic lower-dimensional case, Bourgain solves the problem by a “multiscale theorem,” which he first developed in the context of KAM for PDEs, see Bourgain (2005a) or Berti et al. (2015). Actually as shown in Berti and Bolle (2014) this approach is completely parallel to a KAM scheme. Indeed the existence of a quasi-periodic solution  $\mathbf{i}(\varphi)$  implies the existence of a symplectic change of variables which puts the Hamiltonian in the normal form:

$$H \circ \Phi = \omega(\xi) \cdot y + Q(x, z; \xi) + H^{(\geq 3)} \quad (140)$$

where  $Q$  is a quadratic form in  $z$  which depends on the angles  $x$ .

### Other Chapters in Classical KAM Theory

In this section, we review in a schematic and informal way some developments and applications of KAM theory; for other more exhaustive surveys, we refer to Arnold et al. (2006), Broer et al. (1996, Sect. 6.3), or Sevryuk (2003).

#### 1. Structure of the Kolmogorov set and Whitney smoothness

The Kolmogorov set (i.e., the union of KAM tori), in nearly integrable systems, tends to fill up (in measure) the whole phase space as the strength of the perturbation goes to zero (compare Remark 7-(v) and Remark 10). A natural question is: what is the global geometry of KAM tori?

It turns out that KAM tori smoothly interpolate in the following sense. For  $\varepsilon$  small enough, there exists a  $C^\infty$  symplectic diffeomorphism  $\phi_*$  of the phase space  $\mathcal{M} = B \times \mathbb{T}^d$  of the nearly-integrable, non-degenerate Hamiltonians  $H = K(y) + \varepsilon P(y, x)$  and a Cantor set  $\mathcal{C}^* \subset B$  such that, for each  $y' \in \mathcal{C}^*$ , the set  $\phi_*^{-1}(\{y'\} \times \mathbb{T}^d)$  is a KAM torus for  $H$ ; in other words, the Kolmogorov set is a smooth, symplectic deformation of the fiber bundle  $\mathcal{C}^* \times \mathbb{T}^d$ . Still another way of describing this result is that there exists a smooth function  $K_*: B \rightarrow \mathbb{R}$  such that  $(K + \varepsilon P) \circ \phi_*$  and  $K_*$  agree, together with their derivatives, on  $\mathcal{C}^* \times \mathbb{T}^d$ : we may, thus, say that, in general, nearly integrable

Hamiltonian systems are integrable on Cantor sets of relative big measure.

Functions defined on closed sets which admits  $C^k$  extensions are called *Whitney smooth*; compare (Whitney 1934), where H. Whitney gives a sufficient condition, based on Taylor uniform approximations, for a function to be Whitney  $C^k$ .

The proof of the above result – given, independently, in Chierchia and Gallavotti (1982) and Pöschel (1982) in, respectively, the analytic and the differentiable case – follows easily from the following lemma<sup>40</sup>:

**Lemma 4** *Let  $\mathcal{C} \subset \mathbb{R}^d$  a closed set and let  $\{f_j\}$ ,  $f_0 = 0$ , be a sequence of functions analytic on  $W_j = \cup_{y \in \mathcal{C}} D(y, r_j)$ . Assume that  $\sum_{j \geq 1} \sup_{W_j} |f_j - f_{j-1}| r_j^{-k} < \infty$ . Then,  $f_j$  converges uniformly to a function  $f$ , which is  $C^k$  in the sense of Whitney on  $\mathcal{C}$ .*

Actually, the dependence upon the angles  $x'$  of  $\phi_*$  is analytic and it is only the dependence upon  $y' \in \mathcal{C}^*$  which is Whitney smooth (“anisotropic differentiability,” compare Sect. 2 in Pöschel (1982)).

For more information and a systematic use of Whitney differentiability, see Broer et al. (1996).

#### 2. Power series expansions

KAM tori  $\mathcal{T}_{\omega, \varepsilon} = \phi_\varepsilon(\mathbb{T}^d)$  of nearly integrable Hamiltonians correspond to quasi-periodic trajectories  $z(t; \theta, \varepsilon) = \phi_\varepsilon(\theta + \omega t) = \phi_H^t(z(0; \theta, 0))$ ; compare items (d) and (e) of section “Introduction” and Remark 2-(i) above. While the *actual* existence of such quasi-periodic motions was proven, for the first time, only thanks to KAM theory, the *formal* existence, in terms of formal  $\varepsilon$ -power series<sup>41</sup> was well known in the XIX century to mathematicians and astronomers (such as Newcombe, Lindstedt and, especially, Poincaré; compare (Poincaré), Vol. II). Indeed, formal power solutions of nearly integrable Hamiltonian equations are not difficult to construct (see, e.g., Sect. 7.1 of Celletti and Chierchia (1995)) but *direct proofs* of the convergence of the series, i.e., proofs not based on

Moser's "indirect" argument recalled in Remark 7-(iii) but, rather, based upon direct estimates on the  $k$ th  $\varepsilon$ -expansion coefficient, are quite difficult and were carried out only in the late eighties by H. Eliasson (1996). The difficulty is due to the fact that, in order to prove the convergence of the Taylor-Fourier expansion of such series, one has to recognize compensations among huge terms with different signs<sup>42</sup>. After Eliasson's breakthrough based upon a semi-direct method (compare the "Postscript 1996" at p. 33 of Eliasson (1996)), fully direct proofs were published in 1994 in Chierchia and Falcolini (1994) and Gallavotti (1994).

### 3. Nondegeneracy assumptions

Kolmogorov's nondegeneracy assumption (71) can be generalized in various ways. First of all, Arnold pointed out in Arnold (1963a) that the condition

$$\det \begin{pmatrix} K_{yy} & K_y \\ K_y & 0 \end{pmatrix} \neq 0, \quad (141)$$

(this is a  $(d+1) \times (d+1)$  matrix where last column and last row are given by the  $(d+1)$ -vector  $(K_y, 0)$ ) which is independent from condition (71), is also sufficient to construct KAM tori. Indeed, (141) may be used to construct *iso-energetic* KAM tori, i.e., tori on a *fixed energy level*<sup>43</sup>  $E$ .

More recently, Rüssmann (1989) (see, also, Rüssmann (2001)), using results of Diophantine approximations on manifolds due to Pyartly (1969), formulated the following condition ("Rüssmann non-degeneracy condition"), which is essentially necessary and sufficient for the existence of a positive measure set of KAM tori in nearly integrable Hamiltonian systems: *the image  $\omega(B) \subset \mathbb{R}^d$  of the unperturbed frequency map  $y \rightarrow \omega(y) := K_y(y)$  does not lie in any hyperplane passing through the origin*. We simply add that one of the prices that one has to pay to obtain these beautiful general results is that one cannot fix ahead the frequency.

For a thorough discussion of this topic, see Sect. 2 of Sevryuk (2003).

### 4. Some physical applications

We now mention a short (and non-exhaustive) list of important physical application of KAM theory. For more information, see Sect. 6.3.9 of Arnold et al. (2006) and references therein.

#### 4.1 Perturbation of classical integrable systems

As mentioned above (Remark 1-(iii)), one of the main original motivation of KAM theory is the perturbation theory for nearly integrable Hamiltonian systems. Among the most famous classical integrable systems we recall: one-degree-of freedom systems; Keplerian two-body problem, geodesic motion on ellipsoids; rotations of a heavy rigid body with a fixed point (for special values of the parameters: Euler's, Lagrange's, Kovalevskaya's and Goryachev-Chaplygin's cases); Calogero-Moser's system of particles; see, Sect. 5 of Arnold et al. (2006) and Moser (1983).

A first highly non-trivial step, in order to apply KAM theory to such classical systems, is to construct explicitly action-angle variables and to determine their analyticity properties, which is in itself a technical non-trivial problem. A second problem which arises, especially in Celestial Mechanics, is that the integrable (transformed) Hamiltonian governing the system may be highly degenerate (*proper degeneracies* – see Sect. 6.3.3, B of Arnold et al. (2006)), as is the case of the planetary  $n$ -body problem. Indeed, the first complete proof of the existence of a positive measure set of invariant tori<sup>44</sup> for the planetary  $(n+1)$  problem (one body with mass 1 and  $n$  bodies with masses smaller than  $\varepsilon$ ) has been published only in 2004 (Féjóz 2004) (see, also, Chierchia (2006)); a completion of Arnold's project (1963b) (where Arnold proved the first nontrivial case of the circular planar three body problem and gave a sketch of how to generalize to the general case) has been carried out in

Chierchia and Pinzari (2011a); see also, Chierchia and Pinzari (2010, 2011b, 2014).

#### 4.2 Topological trapping in low dimensions

The general 2-degree-of-freedom nearly integrable Hamiltonian exhibits a kind of stability particularly strong: the phase space is 4-dimensional and the energy levels are 3-dimensional; thus, KAM tori (which are two-dimensional and which are guaranteed, under condition (141), by the iso-energetic KAM theorem) separate the energy levels and orbits lying between two KAM tori will remain forever trapped in such invariant region. In particular the evolution of the action variables stays forever close to the initial position (“total stability”).

This observation is originally due to Arnold (1963a); for applications to the stability of three-body problems in celestial mechanics see Celletti and Chierchia (2007) and item 4.4 below.

In higher dimension, this topological trapping is no more available, and in principle nearby any point in phase space, it may pass an orbit whose action variables undergo a displacement of order one (“Arnold’s diffusion”). A rigorous complete proof of this conjecture is still missing<sup>45</sup>.

#### 4.3 Spectral Theory of Schrödinger operators

KAM methods have been applied also very successfully to the spectral analysis of the one-dimensional Schrödinger (or “Sturm-Liouville”) operator on the real line  $\mathbb{R}$

$$L := -\frac{d^2}{dt^2} + v(t), \quad t \in \mathbb{R}. \quad (142)$$

If the “potential”  $v$  is bounded, then there exists a unique self-adjoint operator on the real Hilbert space  $\mathcal{L}^2(\mathbb{R})$  (the space of Lebesgue square-integrable functions on  $\mathbb{R}$ ) which extends  $L$  above on  $C_0^2$  (the space of twice differentiable functions with compact support). The problem is then to study

the spectrum  $\sigma(L)$  of  $L$ ; for generalities, see Coddington and Levinson (1955).

If  $v$  is periodic, then  $\sigma(L)$  is a continuous band spectrum, as it follows immediately from Floquet theory (Coddington and Levinson 1955). Much more complicate is the situation for quasi-periodic potentials  $v(t) := V(\omega t) = V(\omega_1 t, \dots, \omega_n t)$ , where  $V$  is a (say) real-analytic function on  $\mathbb{T}^n$ , since small-divisor problems appear, and the spectrum can be nowhere dense. For a beautiful classical exposition, see Moser (1983), where, in particular, interesting connections with mechanics are discussed<sup>46</sup>; for deep developments of generalization of Floquet theory (“reducibility”) to quasi-periodic Schrödinger operators, see Avila and Krikorian (2006); Eliasson (1992).

#### 4.4 Physical stability estimates and breakdown thresholds

KAM Theory is perturbative and works if the parameter  $\varepsilon$  measuring the strength of the perturbation is small enough. It is therefore a fundamental question: *how small  $\varepsilon$  has to be in order for KAM results to hold*. The first concrete applications were extremely discouraging: in 1966, the French astronomer M. Hénon (1966) pointed out that Moser’s theorem applied to the restricted three-body problem (i.e., the motion of an asteroid under the gravitational influence of two unperturbed primary bodies revolving on as given Keplerian ellipse) yields existence of invariant tori if the mass ratio of the primaries is less than<sup>47</sup>  $10^{-50}$ . Since then, a lot of progress has been done and, in Celletti and Chierchia (2007), it has been shown via a computer-assisted proof<sup>48</sup>, that, for a restricted-three body model of a subsystem of the Solar system (namely, Sun, Jupiter, and Asteroid Victoria), KAM tori exist for the “actual” physical values (in such model the Jupiter/Sun mass ratio is about  $10^{-3}$ ) and, in this mathematical model – thanks to the trapping mechanism described in item 4.2 above – trap the actual motion of the subsystem.

From a more theoretical point of view, we notice that (compare Remark 2-(ii)) KAM tori (with a fixed Diophantine frequency) are analytic in  $\varepsilon$ ; on the other hand, it is known, at least in lower dimensional settings (such as twist maps), that above a certain critical value KAM tori (curves) cannot exist (Mather 1984). Therefore, there must exist a critical value  $\varepsilon_c(\omega)$  (“breakdown threshold”) such that, for  $0 \leq \varepsilon < \varepsilon_c(\omega)$ , the KAM torus (curve)  $\mathcal{T}_{\omega, \varepsilon}$  exists, while for  $\varepsilon > \varepsilon_c(\omega)$  does not. The mathematical mechanism for the breakdown of KAM tori is far from being understood; for a brief review and references on this topic, see, e.g., Sect. 1.4 in Celletti and Chierchia (2007).

## Infinite Dimensional KAM Theory

One of the most important developments of KAM theory, besides the full applications to classical  $n$ -body problems mentioned above, is the successful extension to infinite dimensional settings, so as to deal with classes of partial differential equations carrying a Hamiltonian or a reversible structure. The concept of integrability for a Hamiltonian PDE has been studied widely since the 1960s. Most of the literature on KAM theory for PDEs however is on the existence of small quasi-periodic solutions for PDEs with an elliptic fixed point at zero and such that the equation linearized at zero has a numerable basis of eigenvectors, either on a *compact manifold* or on  $\mathbb{R}^d$  with a confining potential. Regarding the construction of large quasi-periodic solutions for PDEs close to a nonlinear integrable model, the results are much fewer; see, however, Berti et al. (2018).

It must be remarked that quasi-periodic solutions are the infinite dimensional analogue of lower dimensional tori; hence, they are expected to cover a set of measure zero in phase space. Results on *maximal tori* for PDEs are very few and mostly on ad hoc models; see, e.g., Bourgain (2005b); Chierchia and Perftetti (1995); Pöschel (2002).

The first results on quasi-periodic solutions were obtained by using an adaptation of Theorem

3 (see, for instance, Kuksin (1988), Wayne (1990), Pöschel (1996a)) and were for semi-linear PDEs with Dirichlet boundary conditions in  $[0, \pi]$ . As an example, we state the result for the NLS equation

$$iu_t - u_{xx} + |u|^2 u + f(|u|^2)u \quad (143)$$

**Theorem 5 (Kuksin-Pöschel 1996)** *Suppose the nonlinearity  $f(y)$  is analytic and has a zero of degree at least two in  $y = 0$ . Then for all  $n \in \mathbb{N}$  and all  $S = \{j_1, \dots, j_n\} \subset \mathbb{N}$ , there exists a Cantor manifold  $\varepsilon_S$  of real analytic, linearly stable, diophantine  $n$ -tori for Eq. (143). More precisely there exists a Cantor set  $\mathcal{C}$ , with asymptotically full density at zero, such that for all  $\xi \in \mathcal{C}$  there exists a linearly stable solution of (143) of the form*

$$\begin{aligned} u(t, x; \xi) &= \sum_{j \in S} 2\sqrt{\xi_j} \sin(\omega_j t + jx) \\ &+ o\left(\sqrt{|\xi|}\right), \quad \omega_j := j^2 \\ &+ O(|\xi|), \end{aligned} \quad (144)$$

where  $o(\sqrt{|\xi|})$  is small in some appropriate analytic norm and the map  $\xi \rightarrow u(\xi)$  is Lipschitz continuous.

A first remark is that in this equation there are no parameters; hence, it is not directly written in the setting of Theorem 3. Just as one would do in the finite dimensional case, this problem is overcome by performing a step of Birkhoff normal form, in order to start from an unperturbed system which has a twist, and then using the initial actions as parameters.

In order to concentrate on the problems connected with small divisors, we shall outline the proof only in the simplified case

$$\begin{cases} iu_t - u_{xx} + V * u + |u|^2 u \\ u(t, 0) = u(t, \pi) \end{cases} \quad (145)$$

where  $V * u$  is convolution with an even function  $V = V(x)$ , and we consider its Fourier coefficients  $\{V_j\}_{j \geq 0}$  as parameters.



Just as (143), this is a Hamiltonian system with respect to the symplectic form

$$\omega(u, v) = 2\text{Im} \int_0^\pi u \bar{v}.$$

In order to highlight the equivalence with the problem of lower dimensional tori, we pass this equation in sin-Fourier series

$$u(t, x) := \sqrt{\frac{2}{\pi}} \sum_{j \in \mathbb{N}} u_j(t) \sin(jx)$$

and we write the Hamiltonian as

$$H = \sum_{j \in \mathbb{N}} (j^2 + V_j) |u_j|^2 + \int_0^\pi \left| \sum_{j \in \mathbb{N}} u_j \sin(jx) \right|^4$$

We choose:

- Any finite set  $\mathcal{S} = \{j_1, \dots, j_n\} \subset \mathbb{N}$
- Any initial actions  $\mathcal{I} = \{I_1, \dots, I_n\} \in \mathbb{R}_+$

we fix all the  $V_j$  with  $j \notin \mathcal{S}$  and keep the rest as free parameters. For example, we might fix  $V_j = 0$   $\forall j \notin \mathcal{S}$  and denote  $V_{j_i} = \xi_i$  for  $i = 1, \dots, n$ .

Now we look for small quasi-periodic solutions of the form

$$\begin{aligned} \sqrt{\varepsilon} \sum_{i=1}^n \sqrt{I_i} e^{i\omega_i t} \sin(j_i x) + o(\varepsilon), \\ \omega_i = j_i^2 + \xi_i + O(\varepsilon) \end{aligned} \quad (146)$$

For this purpose, we pass to action-angle variables all the  $\{u_j\}_{j \in \mathcal{S}}$ , by writing

$$\begin{aligned} u_{j_i} &= \sqrt{\varepsilon} \sqrt{I_i + y_i} e^{i x_{j_i}}, i = 1, \dots, n \\ u_j &= \sqrt{\varepsilon} z_j, \quad \forall j \notin \mathcal{S}. \end{aligned}$$

After rescaling the time, the Hamiltonian becomes

$$K^{(0)}(I, \xi) + \sum_{i=1}^n (j_i^2 + \xi_i) y_i + \sum_{j \notin \mathcal{S}} j^2 |z_j|^2 + \varepsilon P(y, x, z),$$

namely, it has the form (132) with  $m = \infty$ . Now in order to apply Theorem 3, we have to specify in

which space the sequence  $\{z_j\}_{j \notin \mathcal{S}}$  lives. Typically one uses a weighted Hilbert space such as

$$\ell_{a,p} := \left\{ z = (z_j)_{j \in \mathbb{N} \setminus \mathcal{S}} : |z|_{a,p}^2 := \sum_{j \in \mathbb{N} \setminus \mathcal{S}} \langle j \rangle^{2p} e^{2a|j|} |z_j|^2 \right\}$$

and redefines the domain  $D(s, r)$  accordingly by substituting  $\mathbb{C}^m$  with  $\ell_{a,p}$ ; see Pöschel and Trubowitz (1987) for an analysis of the properties of analytic functions on a Banach space. One also defines the regular Hamiltonians as those analytic Hamiltonians for which the norm (137) (again substituting  $\mathbb{C}^m$  with  $\ell_{a,p}$ ) is finite. Also in the infinite dimensional case, this class of Hamiltonians is a scale of Banach spaces closed with respect to Poisson brackets, homogeneous projections and which satisfies smoothing estimates for the ultraviolet cut off in the variables  $x$ . Since the proof of Theorem 3 depends only on such properties and is uniform in  $m$ , we have the same result also in this case.

We have proved that for any choice of  $\mathcal{S}$ ,  $\mathcal{I}$  there exists a Cantor-like set  $\Pi^*$  (explicitly defined in (134), and depending on  $\mathcal{S}$ ,  $\mathcal{I}$ ) such that for all  $\xi \in \Pi^*$  there exists quasi-periodic solutions for (145) of the form (146).

One easily verifies that the conditions (136) hold. However, in this infinite dimensional setting this is not enough in order to ensure that the measure of  $\Pi^*$  is positive. By exploiting the fact that  $\Omega_j = j^2 + O(\varepsilon)$ , one can however verify directly that  $|\Pi^*| \sim \gamma$ , provided that  $\tau > n + 1$ .

The same kind of result can be formulated in the more natural case where the potential is multiplicative, and one can prove that for any  $\mathcal{S}$  and for most choices of potential there exist analytic solutions such as (146).

This strategy for proving the existence of finite dimensional invariant tori is quite general and can be applied to many dispersive PDEs on an interval with Dirichlet boundary conditions. A similar strategy can be used also for the Klein-Gordon equation, even though the linear dispersion law makes the measure estimates more complex, see Pöschel (1996b).

This approach, based on applying Theorem 3 in an infinite dimensional setting, has two main drawbacks:



- (A) It relies on the fact that the unperturbed normal frequencies are distinct or at least have finite and uniformly bounded multiplicity as in Chierchia and You (2000). In the context of PDEs, this gives strong restrictions on the domains. For instance, it cannot be applied to equations such as the Nonlinear Schrödinger or the Nonlinear Wave on compact domains without boundary, except in the simplest case of the circle, since the eigenvalues are multiple with unbounded multiplicity.
- (B) By construction the change of coordinates which puts the Hamiltonian in normal form must be the time one flow of a *regular* Hamiltonian, i.e., with finite norm (137). In the infinite dimensional case, this creates unnecessary restrictions, since there exist bounded symplectic changes of variables which are not of this form.

The first results in the direction of removing the assumptions on the multiplicity of eigenvalues were obtained by using a different strategy, proposed by Craig, Wayne for periodic solutions and then developed by Bourgain. This approach is the infinite dimensional analogue of Theorem 5. Actually the first results were in the infinite dimensional setting, and the applications to finite dimensional systems came afterwards. As in the KAM approach, in order to work in an infinite dimensional setting, one needs some knowledge on the asymptotics of the normal sites; we refer to Berti and Bolle (2013), Bourgain (1998, 2005a) or Berti et al. (2015) for details. We remark that these types of results do not imply any stability of the quasi-periodic solutions, nor the existence of a constant coefficients normal form such as the one in (135).

The first results on stable KAM tori on  $\mathbb{T}^d$  were given by Geng and You in 2006, for Nonlinear Wave and Beam equations with a convolution potential. The main ideas were: 1. to exploit the translation invariance of such equations and to use the consequent constants of motion in order to simplify the small divisor problem; 2. to exploit the fact that the nonlinearities in the Wave and Beam equations are 1-*smoothing* in order to prove

the measure estimates (in our notation, this amounts to proving that  $\Pi^*$  has positive Lebesgue measure). The more difficult case of the Nonlinear Schrödinger equation was studied by Eliasson and Kuksin (2009, 2010), where the authors deal with an equation with external parameters, like (145), but with  $x \in \mathbb{T}^d$ . In these papers, the authors do not require translation invariance, instead they deal with clusters of multiple eigenvalues. Moreover, they introduce the notion of *Töplitz-Lipschitz* hamiltonians in order to handle the measure estimates. We mention also the papers (Procesi and Procesi 2015, 2016), which prove existence and stability of quasi-periodic solutions for the NLS equation without outer parameters. The statement of the result is essentially identical to the one of Theorem 3, but there are two main differences:

1. Due to the complicated resonant structure of the NLS on  $\mathbb{T}^d$ , there are some *pathological* choices of tangential sites  $\mathcal{S}$  on which one is not able to prove existence of quasi-periodic solutions and which are the basis of the construction of *weakly turbulent* solutions as in Colliander et al. (2010). More precisely the existence of quasi-periodic solutions is proved for *generic* choices of the tangential sites, i.e., all  $\mathcal{S}$  which are not on the zero set of an explicit but very complicated polynomial.
2. There exist positive measure sets of actions in which solutions exist but there are a finite number of linearly unstable directions.

Concerning results on more complicated manifolds, we mention Berti et al. (2015); Grébert and Paturel (2016) and finally Grébert and Paturel (2011) which deals with a nonlinear quantum harmonic oscillator.

A breakthrough step in overcoming the restrictions explained in point (B) above was first proposed for the much simpler case of periodic solutions in Iooss et al. (2005), in order to study Euler's equations of water waves. This strategy was developed and extended to the quasi-periodic case by Baldi Berti and Montalto, who started by considering an equation of the form

$$u_t + u_{xxx} - 6uu_x - \partial_x[(\partial_u f)(x, u, u_x) - \partial_x((\partial_{u_x} f)(x, u, u_x))] = 0, \quad (147)$$

under periodic boundary conditions  $x \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$ , and assuming that  $f(x, u, v) \in C^q$  has a zero of order at least five in  $u, v = 0$ .

**Theorem 6 (Baldi Berti and Montalto (2015))**

For any  $v \geq 1$  and for all generic choices of tangential sites  $S = \{j_1, \dots, j_v\} \subset \mathbb{N}$ , the KdV equation (147) possesses small amplitude quasi-periodic solutions of the form

$$u(t, x) = \sum_{j \in S} 2\sqrt{\xi_j} \cos(\omega_j t + jx) + o\left(\sqrt{|\xi|}\right), \quad \omega_j := j^3 - 6\xi_{jj}^{-1}, \quad (148)$$

for a “Cantor-like” set of small amplitudes  $\xi \in \mathbb{R}_+^v$  with density 1 at  $\xi = 0$ . The term  $o(\sqrt{|\xi|})$  is small in some  $H^s$ -Sobolev norm,  $s < q$ . These quasi-periodic solutions are linearly stable.

The proof is done by applying a Nash-Moser scheme, as explained in the proof of Theorem 4. The key problem is in inverting a linear unbounded operator  $\mathcal{L}$  of the form

$$\mathcal{L} = \partial_t + (1 + \varepsilon a(x, \omega t)) \partial_x^3 + \varepsilon b(x, \omega t) \partial_x + \varepsilon c(x, \omega t)$$

with  $\omega \in \mathbb{R}^n$  a diophantine vector. The simplest way to invert  $\mathcal{L}$  is to diagonalize it by a bounded change of variables. One could try to construct such change of variables by a KAM scheme: recall that a linear Hamiltonian vector field corresponds to a quadratic Hamiltonian, hence one can try to apply Theorem 3 (putting the Hamiltonian in normal form corresponds to diagonalizing  $\mathcal{L}$ ). This approach however fails, indeed even in the simplest cases it may not be possible to diagonalize  $\mathcal{L}$  by using the flow of a regular Hamiltonian. As an example, assume for simplicity that  $\mathcal{L} = \partial_t + \left(\frac{1}{1+\varepsilon a(x)} \partial_x\right)^3$  where  $a(x)$  has zero mean. Then, clearly, the diagonalizing change of variables is

$$u(x) \rightarrow v(x) = u(x + \varepsilon \beta(x)), \quad \beta_x = a(x),$$

which is not the flow of a regular Hamiltonian but is bounded from  $H^s$  to itself for all  $s$ . In this simple case, the change of variables is constructed by hand directly, in more complicated examples the main feature that one exploits is that  $\mathcal{L}$  is a *pseudo-differential* operator. Then the strategy proposed in Baldi et al. (2015) is:

- (i) Apply changes of variables which are the flow of pseudo-differential vector fields, in order to conjugate  $\mathcal{L}$  to an operator, say  $\widehat{\mathcal{L}}$ , sum of a diagonal operator plus a correction which is a bounded operator of size  $\varepsilon$ .
- (ii) Use a KAM scheme like the one in Theorem 3 in order to diagonalize  $\widehat{\mathcal{L}}$ .

This approach is quite general, and it can be adapted to cover also autonomous equations and has allowed to prove existence and stability for quasi-periodic solutions for many fully nonlinear PDEs on the circle. We mention, among others, the paper (Baldi et al. 2017) where the authors show the existence of quasi-periodic solutions for water waves with gravity.

## Future Directions

Many natural questions, especially in infinite dimensions, remain widely open in KAM theory. In this final section, we briefly mention a (very) few of them.

- (i) In finite dimensional Hamiltonian system, a basic question is still to fully understand the “Kolmogorov set,” i.e., the set of *all* Diophantine invariant maximal tori, including the maximal invariant tori, which are not deformation of integrable tori and which, in general, arise near resonances. In Arnold et al. (2006), it conjectured that the complement of the Kolmogorov set is, in general, bounded by  $\varepsilon$  if  $\varepsilon$  is the size of the perturbation (recall that, as mentioned above, the

complementary of the KAM primary tori – namely, the invariant tori which are deformation of integrable tori – may be bounded by a constant times  $\sqrt{\varepsilon}$ ). In Biasco and Chierchia (2015), it is announced a partial proof of this conjecture in the special case of “mechanical systems,” i.e., Hamiltonian systems of the form  $|p|^2/2 + \varepsilon f(q)$  with  $(p, q) \in \mathbb{R}^d \times \mathbb{T}^d$ .

- (ii) A very interesting and widely open topic in the study of Hamiltonian PDEs, is the study of maximal tori, or possibly even lower dimensional tori of infinite dimension, see, e.g., Bourgain (2005b); Chierchia and Perletti (1995); Pöschel (2002). Such results concern problems with external parameters, of the form say (145). The application to more natural *parameterless* equations, such as (143), is still beyond our reach. In particular it would be interesting to understand the regularity of such almost-periodic solutions and whether they can cover positive measure (in any reasonable sense) sets, as in the finite dimensional case.
- (iii) Another important open problem is that of proving existence of quasi-periodic solutions for general compact Riemannian manifolds. Up to now the, few, results are confined to the case of Zoll manifolds (Grébert and Patrel 2016), or Lie groups (Berti et al. 2015), where there is a very good knowledge of the harmonic analysis.
- (iv) The strategy proposed in Baldi et al. (2015) has allowed to prove existence and stability for many fully nonlinear PDEs on the circle and has been developed, in the similar setting of reducibility, in order to tackle various classes of PDEs in one space variable. Whether this strategy can be generalized in order to cover higher dimensional cases, both on the torus  $\mathbb{T}^d$  or on the line, is a very challenging open problem, in this direction we mention (Bambusi et al. 2018.; Corsi and Montalto 2018).

## Appendix A: The Classical Implicit Function Theorem

Here we discuss the classical Implicit Function Theorem for complex functions from a quantitative point of view. The following Theorem is a simple consequence of the Contraction Lemma, which asserts that a contraction on a closed, nonempty metric space<sup>51</sup> has a unique fixed point, which is obtained as  $\lim_j \rightarrow \infty \Phi^j(u_0)$  for any<sup>52</sup>  $u_0 \in X$ .

### Implicit Function Theorem Let

$$\begin{aligned} F : (y, x) \in D^n(y_0, r) \times D^m(x_0, s) &\subset \mathbb{C}^{n+m} \\ &\rightarrow F(y, x) \in \mathbb{C}^n \end{aligned}$$

be continuous with continuous Jacobian matrix  $F_y$ ; assume that  $F_y(y_0, x_0)$  is invertible and denote by  $T$  its inverse; assume also that

$$\begin{aligned} \sup_{D(y_0, r) \times D(x_0, s)} \|1_n - TF_y(y, x)\| &\leq \frac{1}{2}, \\ \sup_{D(x_0, s)} |F(y_0, x)| &\leq \frac{r}{2\|T\|}. \end{aligned} \quad (149)$$

Then, all solutions  $(y, x) \in D(y_0, r) \times D(x_0, s)$  of  $F(y, x) = 0$  are given by the graph of a unique continuous function  $g: D(x_0, s) \rightarrow D(y_0, r)$  satisfying, in particular,

$$\sup_{D(x_0, s)} |g - y_0| \leq 2\|T\| \sup_{D(x_0, s)} |F(y_0, x)|. \quad (150)$$

*Proof* Let  $X = C(D^m(x_0, s), D^n(y_0, r))$  be the closed ball of continuous function from  $D^m(x_0, s)$  to  $D^n(y_0, r)$  with respect to the supnorm  $\|\cdot\|$  ( $X$  is a nonempty metric space with distance  $d(u, v) = \|u - v\|$ ) and denote  $\Phi(y; x) = y - TF(y, x)$ . Then,  $u \rightarrow \Phi(u) = \Phi(u, \cdot)$  maps  $C(D^m(x_0, s))$  into  $C(D^m(x_0, s))$  and, since  $\partial_y \Phi = 1_n - TF_y(y, x)$ , from the first relation in (149), it follows that  $\Phi$  is a contraction. Furthermore, for any  $u \in C(D^m(x_0, s), D^n(y_0, r))$ ,

$$\begin{aligned}
|\Phi(u) - y_0| &\leq |\Phi(u) - \Phi(y_0)| + |\Phi(y_0) - y_0| \\
&\leq \frac{1}{2} \|u - y_0\| + \|T\| \|F(y_0, x)\| \\
&\leq \frac{1}{2} r + \|T\| \frac{r}{2\|T\|} = r,
\end{aligned}$$

showing that  $\Phi: X \rightarrow X$ . Thus, by the Contraction Lemma, there exists a unique  $g \in X$  such that  $\Phi(g) = g$ , which is equivalent to  $F(g, x) = 0 \forall x$ . If  $F(y_1, x_1) = 0$  for some  $(y_1, x_1) \in D(y_0, r) \times D(x_0, s)$ , it follows that  $|y_1 - g(x_1)| = |\Phi(y_1; x_1) - \Phi(g(x_1), x_1)| \leq \alpha |y_1 - g(x_1)|$ , which implies that  $y_1 = g(x_1)$  and that all solutions of  $F = 0$  in  $D(y_0, r) \times D(x_0, s)$  coincide with the graph of  $g$ . Finally, (150) follows by observing that  $\|g - y_0\| = \|\Phi(g) - y_0\| \leq \|\Phi(g) - \Phi(y_0)\| + \|\Phi(y_0) - y_0\| \leq \frac{1}{2} \|g - y_0\| + \|T\| \|F(y_0, \cdot)\|$ , finishing the proof.

### Additions

- (i) If  $F$  is periodic in  $x$  or/and real on reals, then (by uniqueness) so is  $g$ .
- (ii) If  $F$  is analytic, then so is  $g$  (Weierstrass Theorem, since  $g$  is attained as uniform limit of analytic functions).
- (iii) The factors  $1/2$  appearing in the r.h.s.'s of (149) may be replaced by, respectively,  $\alpha$  and  $\beta$  for any positive  $\alpha$  and  $\beta$  such that  $\alpha + \beta = 1$ .

Taking  $n = m$  and  $F(y, x) = f(y) - x$  for a given  $C^1(D(y_0, r), \mathbb{C}^n)$  function, one obtains the **Inverse Function Theorem** Let  $f: y \in D^n(y_0, r) \rightarrow \mathbb{C}^n$  be a  $C^1$  function with invertible Jacobian  $f_y(y_0)$  and assume that

$$\sup_{D(y_0, r)} \|1_n - T f_y\| \leq \frac{1}{2}, \quad T := f_y(y_0)^{-1}, \quad (151)$$

then there exists a unique  $C^1$  function  $g: D(x_0, s) \rightarrow D(y_0, r)$  with  $x_0 := f(y_0)$  and  $s := r/(2\|T\|)$  such that  $f \circ g(x) = \text{id} = g \circ f$ .

Additions analogous to the above ones holds also in this case.

## Appendix B: Complementary Notes

<sup>1</sup>Actually, the first instance of small divisor problem solved analytically is the linearization of the germs of analytic functions and it due to C.L. Siegel (1942). [Page 5]

<sup>2</sup>The well-known Newton's tangent scheme is an algorithm, which allows to find roots (zeros) of a smooth function  $f$  in a region where the derivative  $f'$  is bounded away from zero. More precisely, if  $x_n$  is an "approximate solution" of  $f(x) = 0$ , i.e.,  $f(x_n) = \varepsilon_n$  is small, then the next approximation provided by Newton's tangent scheme is  $x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}$  [which is the intersection with  $x$ -axis of the tangent to the graph of  $f$  passing through  $(x_n, f(x_n))$ ] and, in view of the definition of  $\varepsilon_n$  and Taylor's formula, one has that  $\varepsilon_{n+1} := f(x_{n+1}) = \frac{1}{2} f''(\xi_n) \varepsilon_n^2 / (f'(x_n))^2$  (for a suitable  $\xi_n$ ) so that  $\varepsilon_{n+1} = O(\varepsilon_n^2) = O(\varepsilon_1^{2^n})$  and, in the iteration,  $x_n$  will converge (at a super-exponential rate) to a root  $\bar{x}$  of  $f$ . This type of extremely fast convergence will be typical in the analysis considered in the present article. [Page 5]

<sup>3</sup>The elements of  $\mathbb{T}^d$  are equivalence classes  $x = \bar{x} + 2\pi\mathbb{Z}^d$  with  $\bar{x} \in \mathbb{R}^d$ . If  $x = \bar{x} + 2\pi\mathbb{Z}^d$  and  $y = \bar{y} + 2\pi\mathbb{Z}^d$  are elements of  $\mathbb{T}^d$ , then their distance  $d(x, y)$  is given by  $\min_{n \in \mathbb{Z}^d} |\bar{x} - \bar{y} + 2\pi n|$  where  $|\cdot|$  denotes the standard Euclidean norm in  $\mathbb{R}^n$ ; a smooth (analytic) function on  $\mathbb{T}^d$  may be viewed as ("identified with") a smooth (analytic) function on  $\mathbb{R}^d$  with period  $2\pi$  in each variable. The torus  $\mathbb{T}^d$  endowed with the above metric is a real-analytic, compact manifold. For more information, see Spivak (1999). [Page 6]

<sup>4</sup>A symplectic form on a (even dimensional) manifold is a closed, nondegenerate differential 2-form. The symplectic form  $\alpha = dy \wedge dx$  is actually *exact* symplectic, meaning that  $\alpha = d(\sum_{i=1}^n y_i dx_i)$ . For general information, see Arnold (1974). [Page 6]

<sup>5</sup>For general facts about the theory of ODE (such as Picard theorem, smooth dependence upon initial data, existence times), see, e.g., Coddington and Levinson (1955). [Page 6]

<sup>6</sup>This terminology is due to the fact that the  $x_j$  are “adimensional” angles, while analyzing the physical dimensions of the quantities appearing in Hamilton’s equations one sees that  $\dim(y) \times \dim(x) = \dim H \times \dim(t)$  so that  $y$  has the dimension of an energy (the Hamiltonian) times the dimension of time, i.e., by definition, the dimension of an action. [Page 7]

<sup>7</sup>This terminology is due to the fact that a classical mechanical systems of  $d$  particles of masses  $m_i > 0$  and subject to a potential  $V(q)$  with  $q \in A \subset \mathbb{R}^d$  is governed by a Hamiltonian of the form  $\sum_{j=1}^d \frac{p_j^2}{2m_i} + V(q)$  and  $d$  may be interpreted as the (minimal) number of coordinates necessary to physically describe the system. [Page 7]

<sup>8</sup>To be precise, Eq. (6) should be written as  $y(t) = v(\pi_{\mathbb{T}^d}(\omega t)), x(t) = \pi_{\mathbb{T}^d}(\omega t + u(\pi_{\mathbb{T}^d}(\omega t)))$  where  $\pi_{\mathbb{T}^d}$  denotes the standard projection of  $\mathbb{R}^d$  onto  $\mathbb{T}^d$ ; however, we normally omit the projection of  $\mathbb{R}^d$  onto  $\mathbb{T}^d$ . [Page 8]

<sup>9</sup>As standard,  $U_\theta$  denotes the  $(d \times d)$  Jacobian matrix with entries  $\frac{\partial U_{ij}}{\partial \theta_j} = \delta_{ij} + \frac{\partial u_i}{\partial \theta_j}$ . [Page 8]

<sup>10</sup>For generalities, see Arnold (1974); in particular, a Lagrangian manifold  $L \subset \mathcal{M}$  which is a graph over  $\mathbb{T}^d$  admits a “generating function,” i.e., there exists a smooth function  $g: \mathbb{T}^d \rightarrow \mathbb{R}$  such that  $L = \{(y, x): y = g_x(x), x \in \mathbb{T}^d\}$ . [Page 8]

<sup>11</sup>Compare Rüssmann (1975) and references therein. We remark that, if  $B(\omega_0, r)$  denote the ball in  $\mathbb{R}^d$  of radius  $r$  centered at  $\omega_0$  and fix  $\tau > d - 1$ , then one can prove that the Lebesgue measure of  $B(\omega_0, r) \setminus \mathcal{D}_{\kappa, \tau}$  can be bounded by  $c_d \kappa r^{d-1}$  for a suitable constant  $c_d$  depending only on  $d$ ; for the simple proof, see, e.g., Chierchia and Peretti (1995). [Page 9]

<sup>12</sup>The sentence “can be put into the form” means “there exists a symplectic diffeomorphism  $\phi: (y, x) \in \mathcal{M} \rightarrow (\eta, \xi) \in \mathcal{M}$  such that  $H \circ \phi$  has the form (10)”; for multiindices  $\alpha, |\alpha| = \alpha_1 + \dots + \alpha_d$  and  $\partial_y^\alpha = \partial_{y_1}^{\alpha_1} \dots \partial_{y_d}^{\alpha_d}$ ; the vanishing of the derivatives of a function  $f(y)$  up to order  $k$  in the origin will also be indicated through the expression  $f = O(|y|^{k+1})$ . [Page 9]

<sup>13</sup>**Notation:** If  $A$  is an open set and  $p \in \mathbb{N}$ , then the  $C^p$ -norm of a function  $f: x \in A \rightarrow f(x)$  is defined as  $\|f\|_{C^p(A)} := \sup_{|\alpha| \leq p} \sup_A |\partial_x^\alpha f|$ . [Page 12]

<sup>14</sup>Standard notation: If  $f$  is a scalar function  $f_y$  is a  $d$ -vector;  $f_{yy}$  the Hessian matrix  $(f_{y_i y_j}); f_{yyy}$  the

symmetric 3-tensor of third derivatives acting as follows:  $f_{yyy} a \cdot b \cdot c := \sum_{i,j,k} \frac{\partial^3 f}{\partial y_i \partial y_j \partial y_k} a_i b_j c_k$ . [Page 14]

<sup>15</sup>Standard notation: If  $f$  is (a regular enough) function over  $\mathbb{T}^d$ , its Fourier coefficients are defined as  $f_n := \int_{\mathbb{T}^d} f(x) e^{-in \cdot x} \frac{dx}{(2\pi)^d}$ ; where, as usual,  $i = \sqrt{-1}$  denotes imaginary unit; for general information about Fourier series see, e.g., Katznelson (2004). [Page 16]

<sup>16</sup>The choice of norms on finite dimensional spaces ( $\mathbb{R}^d, \mathbb{C}^d$ , space of matrices, tensors, etc.) is not particularly relevant for the analysis in this article (since changing norms will change  $d$ -depending constants); however, for matrices, tensors (and, in general, linear operators), it is convenient to work with the “operator norm,” i.e., the norm defined as  $\|L\| = \sup_{u \neq 0} \|Lu\|/\|u\|$ , so that  $\|Lu\| \leq \|L\| \|u\|$ , an estimate, which will be constantly be used; for a general discussion on norms, see, e.g., Kolmogorov and Fomin (1999). [Page 21]

<sup>17</sup>As an example, let us work out the first two estimates, i.e., the estimates on  $\|s_x\|_{\bar{\xi}}$  and  $|b|$ ; actually these estimates will be given on a larger intermediate domain, namely,  $W_{\xi - \frac{\delta}{3}}$ , allowing to give the remaining bounds on the smaller domain  $W_{\bar{\xi}}$  (recall that  $W_s$  denotes the complex domain  $D(0, s) \times \mathbb{T}_s^d$ ). Let  $f(x) := P(0, x) - \langle P(0, \cdot) \rangle$ . By definition of  $\|\cdot\|_{\bar{\xi}}$  and  $M$ , it follows that  $\|f\|_{\bar{\xi}} \leq \|P(0, x)\|_{\bar{\xi}} + \|\langle P(0, \cdot) \rangle\|_{\bar{\xi}} \leq 2M$ . By (P5) with  $p = 1$  and  $\xi' = \xi - \frac{\delta}{3}$ , one gets

$$\|s_x\|_{\xi - \frac{\delta}{3}} \leq \bar{B}_0 \frac{2M}{\kappa} 3^{k_0} \delta^{-k_0},$$

which is of the form (54), provided  $c \geq (\bar{B}_0 2 \cdot 3^{k_0})/\kappa$  and  $v \geq k_0$ . To estimate  $b$ , we need to bound first  $|Q_{yy}(0, x)|$  and  $|P_y(0, x)|$  for real  $x$ . To do this we can use Cauchy estimate: by (P4) with  $p = 2$  and, respectively,  $p = 1$ , and  $\xi' = 0$ , we get

$$\begin{aligned} \|Q_{yy}(0, \cdot)\|_0 &\leq mB_2 C \xi^{-2} \\ &\leq mB_2 C \delta^{-2}, \quad \text{and} \\ \|P_y(0, x)\|_0 &\leq mB_1 M \delta^{-1}, \end{aligned}$$

where  $m = m(d) \geq 1$  is a constant which depend on the choice of the norms, (recall also that  $\delta < \xi$ ). Putting these bounds together, one gets that  $|b|$  can be bounded by the r.h.s. of (54) provided



$c \geq m(B_2 \bar{B}_2 2 \cdot 3^{k_0} + B_1)$ ,  $\mu \geq 1$ ,  $\mu \geq 2$  and  $v \geq k_0 + 2$ . The other bounds in (54) follow easily along the same lines. The factor  $3C$  in front of  $\|\partial_y^2 \tilde{Q}\|_0$  has been inserted to simplify later estimates. [Page 21]

<sup>18</sup>We sketch here the **proof of Lemma 1**. The defining relation  $\psi_\varepsilon \circ \varphi = \text{id}$  implies that  $\alpha(x') = -a(x' + \varepsilon \alpha(x'))$ , where  $\alpha(x')$  is short for  $\alpha(x'; \varepsilon)$  and such equation is a fixed point equation for the nonlinear operator  $f: u \rightarrow f(u) := -a(\text{id} + \varepsilon u)$ . To find a fixed point for this equation, one can use a standard contraction Lemma (see Kolmogorov and Fomin (1999)). Let  $Y$  denote the closed ball (with respect to the supnorm) of continuous functions  $u: \mathbb{T}_{\xi'}^d \rightarrow \mathbb{C}^d$  such that  $\|u\|_{\xi'} \leq L$ . By (55),  $|\text{Im}(x' + \varepsilon u(x'))| < \xi' + \varepsilon_0 L < \xi' + \frac{\delta}{3} = \bar{\xi}$ , for any  $u \in Y$ , and any  $x' \in \mathbb{T}_{\xi'}^d$ , so that  $f: Y \rightarrow Y$ ; notice that, in particular, this means that  $f$  sends periodic functions into periodic functions. Moreover, (55) implies also that  $f$  is a contraction: if  $u, v \in Y$ , then, by the mean value theorem,  $|f(u) - f(v)| \leq L|\varepsilon| |u - v|$  (with a suitable choice of norms), so that, by taking the sup-norm, one has  $\|f(u) - f(v)\|_{\xi'} < \varepsilon_0 L \|u - v\|_{\xi'}$  showing that  $f$  is a contraction. Thus, there exists a unique  $\alpha \in Y$  such that  $f(\alpha) = \alpha$ . Furthermore, recalling that the fixed point is achieved as the uniform limit  $\lim_n \rightarrow \infty f^n(0)$  ( $0 \in Y$ ) and since  $f(0) = -a$  is analytic, so is  $f^n(0)$  for any  $n$  and, hence, by Weierstrass Theorem on the uniform limit of analytic function (see Ahlfors (1978)), the limit  $\alpha$  itself is analytic. In conclusion,  $\varphi \in \mathcal{B}_{\xi'}$  and (56) holds.

Next, for  $(y', x) \in \mathcal{W}_{\bar{\xi}}$ , by (54), one has  $|y' + \varepsilon \beta(y', x)| < \bar{\xi} + \varepsilon_0 L < \bar{\xi} + \frac{\delta}{3} = \xi$  so that (57) holds. Furthermore, since  $\|ea_x\|_{\bar{\xi}} < \varepsilon_0 L < 1/3$  the matrix  $1_d + \varepsilon a_x$  is invertible with inverse given by the “Neumann series”  $(1_d + \varepsilon a_x)^{-1} = 1_d + \sum_{k=1}^{\infty} (-1)^k (\varepsilon a_x)^k = 1_d + \varepsilon S(x; \varepsilon)$ , so that (58) holds. The proof is finished. [Page 22]

<sup>19</sup>From (60) it follows immediately that:  $\langle \partial_y^2 Q_1(0, \cdot) \rangle = \langle \partial_y^2 Q(0, \cdot) \rangle + \varepsilon \langle \partial_y^2 \tilde{Q}(0, \cdot) \rangle = T^{-1}(1_d + \varepsilon T \langle \partial_y^2 \tilde{Q}(0, \cdot) \rangle) = T^{-1}(1_d + \varepsilon R)$  and, in view of (52) and (60), we see that  $\|R\| < L/(2C)$ . Therefore, by (61),  $\varepsilon_0 \|R\| < 1/6 < 1/2$ , implying that  $(1 + \varepsilon R)$  is invertible and

$(1_d + \varepsilon R)^{-1} = 1_d + \sum_{k=1}^{\infty} (-1)^k \varepsilon^k R^k = 1_d + \varepsilon D$  with  $\|D\| \leq \|R\|/(1 - \varepsilon \|R\|) < L/C$ . In conclusion,  $T_1 = (1 + \varepsilon R)^{-1} T = T + \varepsilon D T = T + \varepsilon \tilde{T}$ ,  $\|\tilde{T}\| \leq \|D\| C \leq (L/C) C = L$ . [Page 22]

<sup>20</sup>Actually, there is quite some freedom in choosing the sequence  $\{\xi_j\}$  provided the convergence is not too fast; for general discussion, see, Rüssmann (1980), or, also, Celletti and Chierchia (1987) and Chierchia (1986). [Page 23]

<sup>21</sup>In fact, denoting by  $B_*$  the real  $d$ -ball centered at 0 and of radius  $\theta \xi_*$  for  $\theta \in (0, 1)$ , from Cauchy estimate (48) with  $\xi = \xi_*$  and  $\xi' = \theta \xi_*$ , one has  $\|\phi_* - \text{id}\|_{C^p(B_* \times \mathbb{T}^d)} = \sup_{B_* \times \mathbb{T}^d} \sup_{|\alpha|} |\beta| \leq p \left| \partial_y^\alpha \partial_x^\beta (\phi_* - \text{id}) \right| \leq \sup_{|\alpha|+|\beta| \leq p} \left\| \partial_y^\alpha \partial_x^\beta (\phi_* - \text{id}) \right\|_{\theta \xi_*} \leq B_p \|\phi_* - \text{id}\|_{\xi_*} 1/(\theta \xi_*)^p \leq \text{const}_p |\varepsilon|$  with  $\text{const}_p := B_p DBM 1/(\theta \xi_*)^p$ . An identical estimate hold for  $\|Q_* - Q\|_{C^p(B_* \times \mathbb{T}^d)}$ . [Page 24]

<sup>22</sup>Also in the third millennium, however,  $\varepsilon$ -power expansions turned out to be an important and efficient tool; see Celletti and Chierchia (2007). [Page 24]

<sup>23</sup>A function  $f: A \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is Lipschitz on  $A$  if there exists a constant (“Lipschitz constant”)  $L > 0$  such that  $|f(x) - f(y)| \leq L|x - y|$  for all  $x, y \in A$ . For a general discussion on how Lebesgue measure changes under Lipschitz mappings, see, e.g., Evans and Gariepy (1992). In fact, the dependence of  $\phi_*$  on  $\bar{y}$  is much more regular, compare Remark 11. [Page 25]

<sup>24</sup>In fact, notice that inverse powers of  $\kappa$  appear through (49) (inversion of the operator  $D_\omega$ ); therefore, one sees that the terms in the first line of (54) may be bounded by  $\tilde{c}\kappa^{-2}$  (in defining  $a$  one has to apply the operator  $D_\omega^{-1}$  twice), but then in  $P^{(1)}$  (see (27)) there appears  $\|\beta\|^2$ , so that the constant  $c$  in the second line of (54) has the form (73); since  $\kappa < 1$ , one can replace in (54)  $c$  with  $\hat{c}\kappa^{-4}$  as claimed. [Page 25]

<sup>25</sup>**Proof of Claim C.** Let  $H_0 := H$ ,  $E_0 := E$ ,  $Q_0 := Q$ ,  $K_0 := K$ ,  $P_0 := P$ ,  $\xi_0 := \xi$  and let us assume (inductive hypothesis) that we can iterate  $j$  times Kolmogorov transformation obtaining  $j$  symplectic transformations  $\phi_{i+1}: \mathcal{W}_{\xi_{i+1}} \rightarrow \mathcal{W}_{\xi_i}$ , for  $0 \leq i \leq j-1$ , and  $j$  Hamiltonians  $H_{i+1} = H_i \circ \phi_{i+1} = K_i + \varepsilon^2 P_i$  real-analytic on  $\mathcal{W}_{\xi_i}$  such that

$$\begin{aligned}
& |\omega|, |E_i|, \|Q_i\|_{\xi_i}, \|T_i\| < C, \\
& |\varepsilon|^{2^i} L_i := |\varepsilon|^{2^i} c C^\mu \delta_0^{-\nu} 2^{vi} M_i \leq \frac{\delta_i}{3}, \\
& \forall 0 \leq i \leq j-1. \quad (*)
\end{aligned}$$

By (\*), Kolmogorov iteration (**Step 2**) can be applied to  $H_i$  and therefore all the bounds described in paragraph **Step 2** holds (having replaced  $H, E, \dots, \xi, \delta, H', E', \dots, \xi'$  with, respectively,  $H_i, E_i, \dots, \xi_i, \delta_i, H_{i+1}, E_{i+1}, \dots, \xi_{i+1}$ ); in particular (see (62)) one has, for  $0 \leq i \leq j-1$  (and for any  $|\varepsilon| \leq \varepsilon_0$ ),

$$\begin{aligned}
|E_{i+1}| & \leq |E_i| + |\varepsilon|^{2^i} L_i \|Q_{i+1}\|_{\xi_{i+1}} \\
& \leq \|Q_i\|_{\xi_i} + |\varepsilon|^{2^i} L_i \|\phi_{i+1} - \text{id}\|_{\xi_{i+1}} \\
& \leq |\varepsilon|^{2^i} L_i M_{i+1} \leq M_i L_i \quad (152)
\end{aligned}$$

Observe that the definition of  $D, B$ , and  $L_i$ ,  $|\varepsilon|^{2^i} L_j (3C\delta_j^{-1}) = DB^j |\varepsilon|^{2^j} M_j$ , so that  $L_i < DB^i M_i$ , thus by the second line in (152), for any  $0 \leq i \leq j-1$ ,  $|\varepsilon|^{2^{i+1}} M_{i+1} < DB^i (M_i |\varepsilon|^{2^i})^2$ , which iterated, yields (67) for  $0 \leq i \leq j$ . Next, we show that, thanks to (66), (\*) holds also for  $i = j$  (and this means that Kolmogorov's step can be iterated an infinite number of times). In fact, by (\*) and the definition of  $C$  in (65):  $|E_j| \leq |E| + \sum_{i=0}^{j-1} \varepsilon_0^{2^i} L_i \leq \frac{1}{3C} \sum_{i \geq 0} \delta_i < |E| + \frac{1}{6} \sum_{i \geq 1} 2^{-i} < |E| + 1 < C$ . The bounds for  $\|Q_i\|$  and  $\|T_i\|$  are proven in an identical manner. Now, by (67) <sub>$i=j$</sub>  and (66),  $|\varepsilon|^{2^j} L_j (3C\delta_j^{-1}) = DB^j |\varepsilon|^{2^j} M_j \leq DB^j (DB^j M_j)^{2^j} / (DB^{j+1}) \leq 1/B < 1$ , which implies the second inequality in (\*) with  $i = j$ ; the proof of the induction is finished and one can construct an *infinite sequence* of Kolmogorov transformations satisfying (\*), (152) and (67) for all  $i \geq 0$ . To check (68), we observe that  $|\varepsilon|^{2^i} L_i = \frac{\delta_0}{3C2^i} DB^i |\varepsilon|^{2^i} M_i \leq \frac{1}{2^{i+1}} (|\varepsilon| DBM)^{2^i} \leq \left(\frac{|\varepsilon| DBM}{2}\right)^{i+1}$  and therefore  $\sum_{i \geq 0} |\varepsilon|^{2^i} L_i \leq \sum_{i \geq 1} \left(\frac{|\varepsilon| DBM}{2}\right)^i \leq |\varepsilon| DBM$ . Thus,  $\|Q - Q_*\|_{\xi_*} \leq \sum_{i \geq 0} \|\tilde{Q}_i\|_{\xi_i} \leq |\varepsilon|^{2^i} L_i \leq |\varepsilon| DBM$ ; and analogously for  $|E - E_*|$  and  $\|T - T_*\|$ . To estimate  $\|\phi_* - \text{id}\|_{\xi_*}$ , observe that  $\|\Phi_i - \text{id}\|_{\xi_i} \leq \|\Phi_{i-1} \circ \phi_i - \phi_i\|_{\xi_i} + \|\phi_i - \text{id}\|_{\xi_i} \leq \|\Phi_{i-1} - \text{id}\|_{\xi_{i-1}}$

$+ |\varepsilon|^{2^i} L_i$ , which iterated yields  $\|\Phi_i - \text{id}\|_{\xi_i} \leq \sum_{k=0}^i |\varepsilon|^{2^k} L_k \leq |\varepsilon| DBM$ : taking the limit over  $i$  completes the proof of (68) and the proof of Claim C. [Page 25]

<sup>26</sup>In fact, observe: (i) given any integer vector  $0 \neq n \in \mathbb{Z}^d$  with  $d \geq 2$ , one can find  $0 \neq m \in \mathbb{Z}^d$  such  $n \cdot m = 0$ ; (ii) the set  $\{tn : t > 0 \text{ and } n \in \mathbb{Z}^d\}$  is dense in  $\mathbb{R}^d$ ; (iii) if  $U$  is a neighborhood of  $y_0$ , then  $K_y(U)$  is a neighborhood of  $\omega = K_y(y_0)$ . Thus, by (ii) and (iii), in  $K_y(U)$  there are infinitely many points of the form  $tn$  with  $t > 0$  and  $n \in \mathbb{Z}^d$  to which correspond points  $y(t, n) \in U$  such that  $K_y(y(t, n)) = tn$  and for any of such points one can find, by (i),  $m \in \mathbb{Z}$  such that  $m \cdot n = 0$ , whence  $K_y(y(t, n)) \cdot m = tn \cdot m = 0$ . [Page 28]

<sup>27</sup>This fact was well known to Poincaré, who based on the above argument his nonexistence proof of integral of motions in the general situation; compare Sect. 7.1.1, Arnold et al. (2006). [Page 28]

<sup>28</sup>Compare (91) but observe, that, since  $\hat{P}$  is a trigonometric polynomial, in view of Remark 9- (ii),  $g$  in (97) defines a real-analytic function on  $D(y_0, \bar{r}) \times \mathbb{T}_{\xi'}^d$  with a suitable  $\bar{r} = \bar{r}(\varepsilon)$  and  $\xi' < \xi$ . Clearly is important to see explicitly how the various quantities depend upon  $\varepsilon$ ; this is shortly discussed after Proposition 2. [Page 29]

<sup>29</sup>In fact:  $\|P\|_{r, \xi - \frac{\delta}{2}} \leq M \sum_{|n| > N} e^{-|n|^{\frac{\delta}{2}}} \leq M e^{-\frac{\delta}{4} N}$   
 $\sum_{|n| > N} e^{-|n|^{\frac{\delta}{2}}} \leq M e^{-\frac{\delta}{4} N} \sum_{|n| > 0} e^{-|n|^{\frac{\delta}{2}}} \leq$   
 $\text{const } M e^{-\frac{\delta}{4} N} \delta^{-d} \leq |\varepsilon| M$  if (107) holds and  $N$  is taken as in (105). [Page 30]

<sup>30</sup>Apply the IFT of Appendix A (with  $r$  replaced by  $\bar{r}$ ,  $x_0$  by 0 and  $s$  by  $|\varepsilon|$ ) to  $F(y, \eta) := K_y(y) + \eta \partial_y P_0(y) - K_y(y_0)$  defined on  $D^d(y_0, \bar{r}) \times D^1(0, |\varepsilon|)$ . Using the mean value theorem, Cauchy estimates and (112),  $\|1_d - TF_y\| \leq \|1_d - TK_{yy}\| + |\varepsilon| \|\partial_y^2 P_0\| \leq \|T\| \|K_{yy}\| \bar{r} + \|T\| |\varepsilon| \|\partial_y^2 P_0\| \leq C^2 2 \frac{\bar{r}}{r} + C |\varepsilon| \frac{4}{r^2} M \leq 2C^2 \frac{\bar{r}}{r} + \frac{|\varepsilon| M}{2r\bar{r}} \leq \frac{1}{4} + \frac{1}{16} < \frac{1}{2}$ ; also:  $2\|T\| \|F(y_0, \eta)\| = 2\|T\| \|\eta\| \|\partial_y P_0(y_0)\| < 2C |\varepsilon| M \frac{2}{r} \leq \frac{\varepsilon}{2}$  (where last inequality is due to the second condition in (112)), showing that conditions (149) are fulfilled. Equation (113) comes from (150). Finally, by Cauchy estimates and (113),



$\|A\| \leq C(C_r^2 \cdot \frac{4CM}{r} |\varepsilon| + |\varepsilon| \frac{4}{r^2} M) \leq \frac{12C^3 M}{r^2} |\varepsilon|$  and (115) follows. [Page 32]

<sup>31</sup>Recall note 18 and notice that  $(1_d + A)^{-1} = 1_d + D$  with  $\|D\| \leq \frac{\|A\|}{1-\|A\|} \leq \frac{16}{13} \|A\| < \frac{192}{13} C^3 M |\varepsilon|/r^2$ , where last two inequalities are due to (115). [Page 32]

<sup>32</sup>Lemma 1 can be immediately extended to the  $y'$ -dependent case (which appear as a dummy parameter) as far as the estimates are uniform in  $y'$  (which is the case). [Page 32]

<sup>33</sup>By (119) and (55),  $|\varepsilon| \|g_x\|_{\bar{r}, \bar{\xi}} \leq |\varepsilon| rL \leq r/2$  so that, by (117), if  $y' \in D_{r/2}(y_1)$ , then one has  $y' + \varepsilon g_x(y', \varphi(y', x')) \in D_r(y_0)$ . [Page 33]

<sup>34</sup>The first requirement in (124) is equivalent to require that  $r_0 \leq r$ , which implies that if  $\bar{r}$  is defined as the r.h.s. of (109), then  $\bar{r} \leq r/2$  as required in (111). Next, the first requirement in (112) at the  $(j+1)$ th step of the iteration translates into  $16C^2 r_{j+1}/r_j \leq 1$ , which is satisfied, since, by definition,  $r_{j+1}/r_j = (1/(2\gamma))^{\tau+1} \leq (1/(2\gamma))^2 = 1/(36C^2) < 1/(16C^2)$ . The second condition in (112), which at the  $(j+1)$ th step, reads  $2CM_j r_{j+1}^{-2} |\varepsilon|^{2j}$ , is implied by  $|\varepsilon|^{2j} L_j \leq \delta_j/(3C)$  (corresponding to (55)), which, in turn, is easily controlled along the lines explained in the note 25. [Page 33]

<sup>35</sup>An area-preserving twist mappings of an annulus  $A = [0, 1] \times \mathbb{S}^1$ , ( $\mathbb{S}^1 = \mathbb{T}^1$ ), is a symplectic diffeomorphism  $f = (f_1, f_2): (y, x) \in A \rightarrow f(y, x) \in A$ , leaving invariant the boundary circles of  $A$  and satisfying the twist condition  $\partial_y f_2 > 0$  (i.e.,  $f$  twists clockwise radial segments). The theory of area preserving maps, which was started by Poincaré (who introduced such maps as section of the dynamics of Hamiltonian systems with two degrees of freedom), is, in a sense, the simplest Hamiltonian context. After Poincaré the theory of area-preserving maps became, in itself, a very reach and interesting field of Dynamical Systems leading to very deep and important results due to Herman, Yoccoz, Aubry, Mather, etc.; for generalities and references, see, e.g., Katok and Hasselblatt (1995). [Page 34]

<sup>36</sup>It is not necessary to assume that  $K$  is real-analytic, but it simplifies the exposition. In our case,  $\ell$  is related to the number  $\sigma$  in (69). We recall the definition of Hölder norms: If  $\ell = \ell_0 + \mu$  with  $\ell_0 \in \mathbb{Z}_+$  and  $\mu \in (0, 1)$ , then  $\|f\|_{C^\ell} := \|f\|_{C^{\ell_0}} +$

$\sup_{|x|=\ell_0} \sup_{0 < |x-y| < 1} \frac{|\partial^2 f(x) - \partial^2 f(y)|}{|x-y|^\mu}$ ;  $C^\ell(\mathbb{R}^d)$  denotes the Banach space of functions with finite  $C^\ell$  norm. [Page 34]

<sup>37</sup>To obtain these new estimates, one can first replace  $\theta$  by  $\sqrt{\theta}$  and then use the remark in the note 21 with  $p = 1$ ; clearly the constant  $\sigma$  has to be increased by one unit with respect to the constant  $\sigma$  appearing in (70). [Page 34]

<sup>38</sup>For general references and discussions about Lemma 2 and 3, see, Moser (1966) and Zehnder (1975); an elementary detailed proof can be found, also, in Chierchia (2003). [Page 34]

<sup>39</sup>**Proof of claim M** The first step of the induction consists in constructing  $\Phi_0 = \phi_0$ : this follows from Kolmogorov Theorem (i.e., Remark 7-(i) and Remark 11) with  $\xi = \xi_1 = 1/2$  (assume, for simplicity, that  $Q$  is analytic on  $W_1$  and note that  $\|\varepsilon\|_{\xi_1}^{P^1} \leq C\|\varepsilon\|_{C^0}^P$  by the first inequality in (125)). Now, assume that (129) and (130) holds together with  $C_i < 4C$  and  $\|\partial(\Phi_i - \text{id})\|_{\alpha\xi_{i+1}} < (\sqrt{2} - 1)$  for  $0 \leq i \leq j$  ( $C_0 = C$  and  $C_i$  are as in (65) for, respectively,  $K_0 := K$  and  $K_i$ ). To determine  $\phi_{j+1}$ , observe that, by (129), one has  $H_{j+1} \circ \Phi_{j+1} = (K_{j+1} + \varepsilon P_{j+1}) \circ \phi_{j+1}$  where  $P_j := (P^{j+1} - P^j) \circ \Phi_j$ , which is real-analytic on  $W_{\alpha\xi_{j+1}}$ ; thus we may apply Kolmogorov Theorem to  $K_{j+1} + \varepsilon P_{j+1}$  with  $\xi = \alpha\xi_{j+1}$  and  $\theta = \alpha$ ; in fact, by the second inequality in (125),  $\|P_{j+1}\|_{\alpha\xi_{j+1}} \leq \|P^{j+2} - P^{j+1}\|_{\xi_{j+2}} \leq c\|P\|_{C^\ell}^{\xi_{j+1}}$  and the smallness condition (69) becomes  $|\varepsilon| D_{\xi_{j+1}}^{\ell-\sigma}$  (with  $D := c_* c \|P\|_{C^\ell} (4C)^b 2^{\sigma/2}$ ), which is clearly satisfied for  $|\varepsilon| < D^{-1} \xi^a$ , for some  $a > 0$ . Thus,  $\phi_{j+1}$  has been determined and (notice that  $\alpha^2 \xi_{j+1} = \xi_{j+1}/2 = \xi_{j+2}$ )  $\|\phi_{j+1} - \text{id}\|_{\xi_{j+2}}, \partial(\|\phi_{j+1} - \text{id}\|_{\xi_{j+2}}) \leq |\varepsilon| D_{\xi_{j+1}}^{\xi_{j+1}}$ . Let us now check the domain constraint  $\Phi_j: W_{\alpha\xi_{j+1}} \rightarrow X_{\xi_{j+1}}$ . By the inductive assumptions and the real-analyticity of  $\Phi_j$ , one has that, for  $z \in W_{\alpha\xi_{j+1}}$ ,  $|\text{Im}\Phi_j(z)| = |\text{Im}(\Phi_j(z) - \Phi_j(\text{Re}z))| \leq |\Phi_j(z) - \Phi_j(\text{Re}z)| \leq \|\partial\Phi_j\|_{\alpha\xi_{j+1}} |\text{Im}z| \leq (1 + \|\partial(\Phi_{j+1} - \text{id})\|_{\alpha\xi_{j+2}}) \alpha\xi_{j+1} < \sqrt{2} \alpha\xi_{j+1} = \xi_{j+1}$  so that  $\Phi_j: W_{\alpha\xi_{j+1}} \rightarrow X_{\xi_{j+1}}$ . The remaining inductive assumptions in (130) with  $j$  replaced by  $j+1$  are easily checked by arguments similar to those used in the induction proof of Claim C above. [Page 35]

<sup>40</sup>See, e.g., the Proposition at page 58 of Chierchia (1986) with  $g_j = f_j - f_{j-1}$ . In fact, the lemma applies to the Hamiltonians  $H_j$  and to the symplectic map  $\phi_j$  in (83) in Arnold's scheme with  $W_j$  in (82) and taking  $\mathcal{C} = \mathcal{C}_* := \left\{ y' = \lim_{j \rightarrow \infty} y_j(\omega) : \omega \in B \cap K_y^{-1}(\mathcal{D}_{\kappa, \tau}) \right\}$  and  $y_j(\omega) := y_j$  is as in (83). [Page 41]

<sup>41</sup>A formal  $\varepsilon$ -power series quasi-periodic trajectory, with rationally independent frequency  $\omega$ , for a nearly integrable Hamiltonian  $H(y, x; \varepsilon) := K(y) + \varepsilon P(y, x)$  is, by definition, a sequence of functions  $\{z_k\} := (\{v_k\}, \{u_k\})$ , real-analytic on  $\mathbb{T}^d$  and such that  $D_\omega z_k = J_{2d} \pi_k \left( \nabla H \left( \sum_{j=0}^{k-1} \varepsilon^j z_j \right) \right)$  where  $\pi_k(\cdot) := \frac{1}{k!} \partial_\varepsilon^k(\cdot)|_{\varepsilon=0}$ ; compare Remark 1-(ii) above. [Page 42]

<sup>42</sup>In fact, Poincaré was not at all convinced of the convergence of such series: see Chap. XIII, n° 149, entitled “Divergence des séries de M. Lindstedt,” of his book (Poincaré). [Page 42]

<sup>43</sup>(71) guarantees that the map from  $y$  in the  $(d-1)$ -dimensional manifold  $\{K = E\}$  to the  $(d-1)$ -dimensional real projective space  $\{\omega_1 : \omega_2 : \dots : \omega_d\} \subset \mathbb{RP}^{d-1}$  (where  $\omega_i = K_{y_i}$ ) is a diffeomorphism. For a detailed proof of the “iso-energetic KAM Theorem,” see, e.g., Delshams and Gutierrez (1996). [Page 43]

<sup>44</sup>Actually, it is not known if such tori are KAM tori in the sense of the definitions given above! [Page 43]

<sup>45</sup>The first example of a nearly integrable system (with two parameters) exhibiting Arnold's diffusion (in a certain region of phase space) was given by Arnold in 1964; a theory for “a priori unstable systems” (i.e., the case in which the integrable system carries also a partially hyperbolic structure) has been worked out in Chierchia and Gallavotti (1994) and in recent years a lot of literature has been devoted to study the “a priori unstable” case and to try to attack the general problem (see, e.g., Sect. 6.3.4 of Arnold et al. (2006) for a discussion and further references). We mention that J. Mather has recently announced a complete proof of the conjecture in a general case Mather (2003). [Page 44]

<sup>46</sup>Here, we mention briefly a different and very elementary connection with classical mechanics. To study the spectrum  $\sigma(L)$  ( $L$  as

above with a quasi-periodic potential  $V(\omega_1 t, \dots, \omega_n t)$ ) one looks at the equation  $\ddot{q} = (V(\omega t) - \lambda)q$ , which is the  $q$ -flow of the Hamiltonian  $\phi_H^t H = H(p, q, I, \varphi; \lambda) := \frac{p^2}{2} + [\lambda - V(\varphi)] \frac{q^2}{2}$  where  $(p, q) \in \mathbb{R}^2$  and  $(I, \varphi) \in \mathbb{R}^n \times \mathbb{T}^n$  (with respect to the standard form  $dp \wedge dq + dI \wedge d\varphi$ ) and  $\lambda$  is regarded as a parameter. Notice that  $\dot{\varphi} = \omega$  so that  $\varphi = \varphi_0 + \omega t$  and that the  $(p, q)$  decouples from the  $I$ -flow, which is, then, trivially determined once the  $(p, q)$  flow is known. Now, the action-angle variables  $(J, \theta)$  for the harmonic oscillator  $\frac{p^2}{2} + \lambda \frac{q^2}{2}$  are given by  $J = r^2 / \sqrt{\lambda}$  and  $(r, \theta)$  are polar coordinates in the  $(p, \sqrt{\lambda} q)$ -plane; in such variables,  $H$  takes the form  $H = \omega \cdot I + \sqrt{\lambda} J - \frac{V(\varphi)}{\sqrt{\lambda}} \sin^2 \theta$ . Now, if, for example  $V$  is small, this Hamiltonian is seen to be a perturbation of  $(n+1)$  harmonic oscillator with frequencies  $(\omega, \sqrt{\lambda})$  and it is remarkable that one can provide a KAM scheme, which preserves the linear-in-action structure of this Hamiltonian and selects the (Cantor) set of values of the frequency  $\alpha = \sqrt{\lambda}$  for which the KAM scheme can be carried out so as to conjugate  $H$  to a Hamiltonian of the form  $\omega \cdot I + \alpha J$ , proving the existence of (generalized) quasi-periodic eigenfunctions. For more details along these lines, see Chierchia (1986). [Page 44]

<sup>47</sup>The value  $10^{-50}$  is about the proton-Sun mass ratio: the mass of the Sun is about  $1.991 \cdot 10^{30}$  Kg, while the mass of a proton is about  $1.672 \cdot 10^{-21}$  Kg, so that (mass of a proton)/(mass of the Sun)  $\simeq 8.4 \cdot 10^{-52}$ . [Page 45]

<sup>48</sup>“Computer-assisted proofs” are mathematical proofs, which use the computers to give rigorous upper and lower bounds on chains of long calculations by means of the so-called “interval arithmetic”; see, e.g., Appendix C of Celletti and Chierchia (2007) and references therein. [Page 45]

<sup>49</sup>Simple examples of such orbits are equilibria and periodic orbits: in such cases there are no small-divisor problems and existence was already established by Poincaré by means of the standard Implicit Function Theorem; see Poincaré, Vol. I, Chap. III. [Page 36]

<sup>50</sup>Typically,  $\xi$  may indicate an initial datum  $y_0$  and  $y$  the distance from such point or (equivalently, if the system is nondegenerate in

the classical Kolmogorov sense)  $\xi \rightarrow \omega(\xi)$  might be simply the identity, which amounts to consider the unperturbed frequencies as parameter; the approach followed here is that in Pöschel (1996a), where, most interestingly,  $m$  is allowed to be  $\infty$ . [Page 36]

<sup>51</sup>I.e., a map  $\Phi: X \rightarrow X$  for which  $\exists 0 < \alpha < 1$  such that  $d(\Phi(u), \Phi(v)) \leq \alpha d(u, v)$ ,  $\forall u, v \in X$ ,  $d(\cdot, \cdot)$  denoting the metric on  $X$ ; for generalities on metric spaces, see, e.g., Kolmogorov and Fomin (1999). [Page 51]

<sup>52</sup> $\Phi^j = \Phi \circ \dots \circ \Phi$   $j$ -times. In fact, let  $u_j := \Phi^j(u_0)$  and notice that, for each  $j \geq 1$   $d(u_{j+1}, u_j) \leq \alpha d(u_j, u_{j-1}) \leq \alpha^j d(u_1, u_0) =: \alpha^j \beta$ , so that, for each  $j, h \geq 1$ ,  $d(u_{j+h}, u_j) \leq \sum_{i=0}^{h-1} d(u_{j+i+1}, u_{j+i}) \leq \sum_{i=0}^{h-1} \alpha^{j+i} \beta \leq \alpha^{j+1} \beta / (1 - \alpha)$ , showing that  $\{u_j\}$  is a Cauchy sequence. Uniqueness is obvious. [Page 51]

## Bibliography

- Ahlfors LV (1978) Complex analysis. An introduction to the theory of analytic functions of one complex variable. In: International series in pure and applied mathematics, 3rd edn. McGraw-Hill Book Co, New York, xi+331 pp
- Arnold VI (1963a) Proof of a Theorem by A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian. *Russ Math Surv* 18:13–40
- Arnold VI (1963b) Small divisor problems in classical and celestial mechanics. *Russ Math Surv* 18:85–191
- Arnold VI (1964) Instability of dynamical systems with many degrees of freedom. (Russian). *Dokl Akad Nauk SSSR* 156:9–12
- Arnold VI (1974) Mathematical methods of classical mechanics. Series: Graduate texts in mathematics, vol 60. Originally published by Nauka, Moscow, 1974 2nd ed. 1989. Corr. 4th printing, 1997, XVI, 509
- Arnold VI, Kozlov VV, Neishtadt AI (eds) (2006) Mathematical aspects of classical and celestial mechanics, dynamical systems III, Series: Encyclopaedia of mathematical sciences, vol 3, 3rd edn. Springer-Verlag, Berlin
- Avila A, Krikorian R (2006) Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles. *Ann Math* 164:911–940
- Baldi P, Berti M, Montalto R (2014) KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation. *Math Ann* 359:471–536
- Baldi P, Berti M, Montalto R (2015) KAM for autonomous quasi-linear perturbations of KdV. *Ann. I. H. Poincaré (C) Anal. Non Linéaire*
- Baldi P, Berti M, Haus E, Montalto R (2017) Time quasi-periodic gravity water waves in finite depth, preprint, arXiv:1708.01517
- Bambusi D, Grebert B, Maspero A, Robert D (2018) Reducibility of the quantum harmonic oscillator in  $d$ -dimensions with polynomial time dependent perturbation. *Anal. PDE* 11(3):775–799
- Berti M, Biasco L (2011) Branching of cantor manifolds of elliptic tori and applications to PDEs. *Commun Math Phys* 305(3):741–796
- Berti M, Bolle P (2013) Quasi-periodic solutions with Sobolev regularity of NLS on  $\mathbb{T}^d$  with a multiplicative potential. *Eur J Math* 15:229–286
- Berti M, Bolle P (2014) A Nash-Moser approach to KAM theory, fields institute communications, vol 75. In: Hamiltonian PDEs and applications, pp 255–284
- Berti M, Montalto R (2016) Quasi-periodic standing wave solutions of gravity-capillary water waves, preprint arXiv:1602.02411 (to appear in *Memoirs of the AMS*, MEMO 891)
- Berti M, Biasco L, Procesi M (2014) KAM theory for the reversible derivative wave equation. *Arch Ration Mech Anal* 212:905–955
- Berti M, Corsi L, Procesi M (2015) An abstract Nash-Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous spaces. *Commun Math Phys* 334(3):1413–1454
- Berti M, Kappeler T, Montalto R (2018) Large KAM tori for perturbations of the dNLS equation, preprint arXiv:1603.09252 (to appear on *Astérisque*) 403
- Biasco L, Chierchia L (2015) On the measure of Lagrangian invariant tori in nearly-integrable mechanical systems. *Atti Accad Naz Lincei Rend Lincei Math Appl* 26(4):423–432
- Biasco L, Chierchia L (2018) Explicit estimates on the measure of primary KAM tori. *Ann Mat Pura Appl* (4) 197(1):261–281
- Bourgain J (1994) Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE. *Int Math Res Not* 1994(11):475–497
- Bourgain J (1996) Gibbs measures and quasi-periodic solutions for nonlinear Hamiltonian partial differential equations. In: Gelfand mathematical seminars. Birkhäuser Boston, Boston, pp 23–43
- Bourgain J (1997) On Melnikov's persistency problem. *Math Res Lett* 4(4):445–458
- Bourgain J (1998) Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. *Ann Math* 148:363–439
- Bourgain J (1999) Periodic solutions of nonlinear wave equations, Harmonic analysis and partial differential equations. Chicago Lectures in mathematics, University of Chicago Press, Chicago, New York, pp 69–97
- Bourgain J (2005a) Green's function estimates for lattice Schrödinger operators and applications, *Annals of mathematics studies*, vol 158. Princeton University Press, Princeton
- Bourgain J (2005b) On invariant tori of full dimension for the 1d periodic NLS. *J Funct Anal* 229:62–94

- Broer HW (1954) KAM theory: the legacy of A. N. Kolmogorov's 1954 paper. *Bull Amer. Math Soc. (N.S.)* 41 (2004), no. 4:507–521
- Broer HW, Huitema GB, Sevryuk MB (1996) Quasi-periodic motions in families of dynamical systems. Order amidst chaos. Lecture notes in mathematics, 1645. Springer-Verlag, Berlin
- Celletti A, Chierchia L (1987) Rigorous estimates for a computer-assisted KAM theory. *J Math Phys* 28:2078–2086
- Celletti A, Chierchia L (1988) Construction of analytic KAM surfaces and effective stability bounds. *Commun Math Phys* 118:119–161
- Celletti A, Chierchia L (1995) A constructive theory of lagrangian tori and computer-assisted applications. In: Jones CKRT, Kirchgraber U, Walther HO (Managing Editors) Dynamics reported. Springer-Verlag, Berlin/London, 4 (New Series), pp 60–129
- Celletti A, Chierchia L (2007) KAM stability and celestial mechanics. *Mem Amer Math Soc* 187(878), viii+134 pp
- Chierchia L (1986) Quasi-periodic Schrödinger operators in one dimension, Absolutely continuous spectra, Bloch waves and integrable Hamiltonian systems. “Quaderni del Consiglio Nazionale delle Ricerche”, (Firenze, Italia), 127 pp. Download: <http://www.mat.uniroma3.it/users/chierchia/REPRINTS/CNR86.pdf>
- Chierchia L (2003) KAM lectures. In: Dynamical systems. Part I: Hamiltonian systems and celestial mechanics. Pubblicazioni della Classe di Scienze, Scuola Normale Superiore, Pisa, pp 1–56, Centro di Ricerca Matematica “Ennio De Giorgi”: Proceedings
- Chierchia L (2006) KAM theory and celestial mechanics. In: Franoise J-P, Naber GL, Tsou ST (eds) Encyclopedia of mathematical physics, vol 3. Elsevier, Oxford, pp 189–199, ISBN 978-0-1251-2666-3
- Chierchia L (2008) A. N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. A comment on: “On conservation of conditionally periodic motions for a small change in Hamilton's function” [Dokl Akad Nauk SSSR (N.S.) 98(1954):527–530; MR0068687]. *Regul Chaotic Dyn* 13(2):130–139
- Chierchia L, Falcolini C (1994) A direct proof of a theorem by Kolmogorov in Hamiltonian systems. *Ann Sc Norm Su Pisa Serie IV* 21:541–593
- Chierchia L, Gallavotti G (1982) Smooth prime integrals for quasi-integrable Hamiltonian systems. *Il Nuovo Cimento* 67 B:277–295
- Chierchia L, Gallavotti G (1994) Drift and diffusion in phase space. *Ann Inst Henri Poincaré Phys Théor* 60:1–144, see also *Erratum*, *Ann Inst Henri Poincaré, Phys Théor*, 68, no. 1:135 (1998)
- Chierchia L, Perfetti P (1995) Second order Hamiltonian equations on  $\mathbb{T}^\infty$  and almost-periodic solutions. *J Differ Equ* 116(1):172–201
- Chierchia L, Pinzari G (2010) Properly-degenerate KAM theory (following V. I. Arnold). *Discrete Contin Dyn Syst Ser S* 3(4):545–578
- Chierchia L, Pinzari G (2011a) The planetary N-body problem: symplectic foliation, reductions and invariant tori. *Invent Math* 186(1):177
- Chierchia L, Pinzari G (2011b) Planetary Birkhoff normal forms. *J Mod Dyn* 5(4):623–664
- Chierchia L, Pinzari G (2014) Metric stability of the planetary N-body problem. In: Proceedings of the international congress of mathematicians–Seoul 2014, vol III. Kyung Moon Sa, Seoul, pp 547–570
- Chierchia L, Qian D (2004) Moser's theorem for lower dimensional tori. *J Differ Equ* 206:55–93
- Chierchia L, You J (2000) KAM Tori for 1D nonlinear wave equations with periodic boundary conditions. *Commun Math Phys* 211:497–525
- Coddington EA, Levinson N (1955) Theory of ordinary differential equations. McGraw-Hill, New York, 14+429 pp
- Colliander J, Keel M, Staffilani G, Takaoka H, Tao T (2010) Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. *Invent Math* 181(1):39–113
- Corsi L, Montalto R (2018) Quasi-periodic solutions for the forced Kirchhoff equation on  $\mathbb{T}^d$ . arXiv: 1802.04139
- Delshams A, Gutierrez P (1996) Effective stability and KAM theory. *J Differ Equ* 128(2):415–490
- Eliasson LH (1988) Perturbations of stable invariant tori for Hamiltonian systems. *Ann Scuola Norm Sup Pisa Cl Sci* 15:115–147
- Eliasson LH (1992) Floquet solutions for the 1-dimensional quasi-periodic Schrödinger equation. *Commun Math Phys* 146:447–482
- Eliasson LH (1996) Absolutely convergent series expansion for quasi-periodic motions. *Math Phys Electron J* 2(Paper 4):33, (electronic); see also: Report 2–88, Department of Mathematics, University of Stockholm (1988)
- Eliasson LH, Kuksin SB (2009) On reducibility of Schrödinger equations with quasiperiodic in time potentials. *Commun Math Phys* 286:125–135
- Eliasson LH, Kuksin SB (2010) KAM for non-linear Schrödinger equation. *Ann Math* 172:371–435
- Eliasson LH, Grebert B, Kuksin SB (2016) KAM for the nonlinear beam equation. *Geom Funct Anal* 26(6):1588–1715
- Evans LC, Gariepy RF (1992) Measure theory and fine properties of functions, Studies in advanced mathematics. CRC Press, Boca Raton, viii+268 pp
- Féjoz J (2004) Démonstration du ‘théorème d’Arnold sur la stabilité du système planétaire (d’après Herman). (French). *Ergodic Theory Dynam Systems* 24(5):1521–1582
- Gallavotti G (1994) Twistless KAM tori, quasi flat homoclinic intersections and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review *Rev Math Phys* 6:343–411

- Geng J, You J (2006) A KAM theorem for Hamiltonian partial differential equations in higher dimensional spaces. *Commun Math Phys* 262(2):343–372
- Geng J, Xu X, You J (2011) An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation. *Adv Math* 226:5361–5402
- Graff S (1974) On the continuation of stable invariant tori for Hamiltonian systems. *J Differ Equ* 15:1–69
- Grébert B, Kappeler T (2014) The defocusing NLS equation and its normal form. *European Mathematical Society (EMS), Zurich*
- Grébert B, Paturel E (2011) KAM for the quantum harmonic oscillator. *Commun Math Phys* 307(2):383–427
- Grébert B, Paturel E (2016) KAM for the Klein Gordon equation on  $\mathbb{S}^d$ . *Boll Unione Mat Ital* 9(2):237–288
- Hénon M (1966) Explorations numériques du problème restreint IV: Masses égales, orbites non périodique. *Bulletin Astronomique* 3., 1, fasc. 2:49–66
- Iooss G, Plotnikov PI, Toland JF (2005) Standing waves on an infinitely deep perfect fluid under gravity. *Arch Ration Mech Anal* 177(3):367–478
- Kappeler T, Pöschel J (2003) *KAM and KdV*. Springer, Berlin
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: *Encyclopedia of mathematics and its applications*, vol 54. New York, Cambridge University Press, xviii+802 pp
- Katzenelson Y (2004) An introduction to harmonic analysis, 3rd edn. Cambridge Mathematical Library/Cambridge University Press, Cambridge, MA, xviii+314 pp
- Kolmogorov AN (1954) On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian. *Dokl Akad Nauk SSR* 98:527–530
- Kolmogorov AN, Fomin SV (1999) *Elements of the theory of functions and functional analysis*. Courier Dover Publications, Mineola, p 288
- Kuksin SB (1987) Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum. *Funktsional Anal i Prilozhen* 21(3):22–37, 95
- Kuksin SB (1988) Perturbation theory of conditionally periodic solutions of infinite-dimensional Hamiltonian systems and its applications to the Korteweg-de Vries equation. (Russian) *Mat Sb (N.S.)* 136(178):396–412, 431, no. 3; translation in *Math USSR-Sb.* 64(1989):397–413, no. 2
- Kuksin SB (1998) A KAM theorem for equations of the Korteweg-de Vries type. *Rev Math Phys* 10(3):1–64
- Kuksin SB (2000) *Analysis of Hamiltonian PDEs*. Oxford lecture series in mathematics and its applications, 19. Oxford University Press, Oxford
- Kuksin SB (2004) Fifteen years of KAM for PDE. *Geometry, topology, and mathematical physics*. *Amer Math Soc Transl Ser* 212(2):237–258, Amer Math Soc, Providence, R. I
- Kuksin SB, Pöschel J (1996) Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. *Ann Math* 2(143):149–179
- Liu J, Yuan X (2011) A KAM theorem for Hamiltonian partial differential equations with unbounded perturbations. *Commun Math Phys* 307(3):629–673
- Mather JN (1984) Nonexistence of invariant circles. *Erg Theory Dynam Syst* 4:301–309
- Mather JN (2003) Arnold diffusion. I. Announcement of results. (Russian). *Sovrem Mat Fundam Napravl* 2:116–130, (electronic); translation in *J. Math. Sci. (N. Y.)* 124 (2004):5275–5289, no. 5
- Melnikov VK (1965) On certain cases of conservation of almost periodic motions with a small change of the Hamiltonian function. (Russian). *Dokl Akad Nauk SSSR* 165:1245–1248
- Moser JK (1961) A new technique for the construction of solutions of nonlinear differential equations. *Proc Natl Acad Sci U S A* 47:1824–1831
- Moser JK (1962) On invariant curves of area-preserving mappings of an annulus. *Nach Akad Wiss Göttingen Math Phys Kl II* 1:1–20
- Moser JK (1966) A rapidly convergent iteration method and non-linear partial differential equations. *Ann Scuola Norm Sup Pisa* 20:499–535
- Moser JK (1967) Convergent series expansions for quasi-periodic motions. *Math Ann* 169:136–176
- Moser JK (1970) On the construction of almost periodic solutions for ordinary differential equations. In: *Proceedings of the international conference on functional analysis and related topics (Tokyo, 1969)*, University of Tokyo Press, Tokyo, pp 60–67
- Moser JK (1983) *Integrable Hamiltonian systems and spectral theory, Lezioni Fermiane. [Fermi lectures]*. Scuola Normale Superiore, Pisa, iv+85 pp
- Nash J (1956) The imbedding problem for Riemannian manifolds. *Ann Math* 63(2):20–63
- Poincaré H *Les méthodes nouvelles de la Mécanique céleste*, tome 1, Paris, Gauthier-Villars, 1892, 385 pp; tome 2, Paris, Gauthier-Villars, 1893, VIII + 479 pp; tome 3, Paris, Gauthier-Villars, 1899, 414 pp
- Pöschel J (1982) Integrability of Hamiltonian systems on cantor sets. *Commun Pure Appl Math* 35:653–695
- Pöschel J (1989) On elliptic lower dimensional Tori in Hamiltonian systems. *Math Z* 202:559–608
- Pöschel J (1996a) A KAM-theorem for some nonlinear PDEs. *Ann Scuola Norm Sup Pisa Cl Sci* 23:119–148
- Pöschel J (1996b) Quasi-periodic solutions for a nonlinear wave equation. *Comment Math Helv* 71(2):269–296
- Pöschel J (2002) On the construction of almost periodic solutions for a nonlinear Schrödinger equation. *Ergodic Theory Dynam Syst* 22(5):1537–1549
- Pöschel J, Trubowitz E (1987) *Inverse spectral theory*. Academic Press, Boston
- Procesi C, Procesi M (2015) A KAM algorithm for the non-linear Schrödinger equation. *Adv Math* 272:399–470
- Procesi C, Procesi M (2016) Reducible quasi-periodic solutions for the non linear Schrödinger equation. *BUMI* 9(2):189–236



- Pyartli AS (1969) Diophantine approximations of submanifolds of a Euclidean space. *Funktsional Anal i Prilozhen* 3(4):59–62, (Russian). English translation in: *Funct Anal Appl* 3(1969):303–306
- Rüssmann H (1975) On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In: Moser J (ed) *Dynamical systems, theory and applications*, Lecture notes in physics, vol 38. Springer, Berlin/Heidelberg, pp 598–624
- Rüssmann H (1976) Note on sums containing small divisors. *Commun Pure Appl Math* 29(6):755–758
- Rüssmann H (1980) On the one-dimensional Schrödinger equation with a quasiperiodic potential. *Nonlinear dynamics* (International Conference, New York, 1979). *Ann N Y Acad Sci* 357:90–107. New York Acad Sci, New York
- Rüssmann H (1989) Nondegeneracy in the perturbation theory of integrable dynamical systems, Number theory and dynamical systems (York, 1987). London mathematical society lecture note series, vol 134. Cambridge University Press, Cambridge, MA, pp 5–18
- Rüssmann H (2001) Invariant tori in non-degenerate nearly integrable Hamiltonian systems. *Regul Chaotic Dyn* 6(2):119–204
- Rüssmann H (1970) Kleine Nenner I Über invariante Kurven differenzierbarer Abbildungen eines Kreisinges (German). *Nachr Akad Wiss Göttingen Math-Phys Kl II* 1970:67–105
- Salamon D (2004) The Kolmogorov-Arnold-Moser theorem. *Math Phys Electron J* 10(Paper 3):37 (electronic)
- Sevryuk MB (2003) The classical KAM theory and the dawn of the twenty-first century. *Moscow Math J* 3(3):1113–1144
- Siegel CL (1942) Iteration of analytic functions. *Ann Math* 43:607–612
- Spivak M (1999) A comprehensive introduction to differential geometry, vol 1, 3rd edn. Publish or Perish, Inc, Houston
- Wayne CE (1990) Periodic and quasi-periodic solutions for nonlinear wave equations via KAM theory. *Commun Math Phys* 127:479–528
- Whitney H (1934) Analytic extensions of differentiable functions defined in closed sets. *Trans Am Math Soc* 36(1):63–89
- Zakharov et al (1991) What is integrability? Springer series in Nonlinear dynamics. Springer-Verlag
- Zehnder E (1975) Generalized implicit function theorems with applications to some small divisor problems. I. *Commun Pure Appl Math* 28:91–14
- Zehnder E (1976) Generalized implicit function theorems with applications to some small divisor problems. II. *Commun Pure Appl Math* 29(1):49–111
- edition published by VINITI, Moscow, 2nd expanded and rev. ed. 2001
- Berretti A, Celletti A, Chierchia L, Falcolini C (1992) Natural boundaries for area-preserving twist maps. *J Stat Phys* 66:1613–1630
- Bost JB (1984/1985) Tores invariants des systèmes dynamiques hamiltoniens, Séminaire Bourbaki, exp. 639
- Broer HW, Sevryuk MB (2010) KAM theory: quasi-periodicity in dynamical systems. In: *Handbook of dynamical systems*. Elsevier, 249–344
- Celletti A, Chierchia L (2006) KAM tori for N-body problems: a brief history. *Celest Mech Dyn Astron* 95(1–4):117–139
- Chierchia L, Falcolini C (1996) Compensations in small divisor problems. *Commun Math Phys* 175:135–160
- de la Llave R (2001) A tutorial on KAM theory, Smooth ergodic theory and its applications (Seattle, WA, 1999). In: *Proceedings of symposia in pure mathematics*, vol 69. American Mathematical Society, Providence, pp 175–292
- de la Llave R, Rana D (1990) Accurate strategies for small divisor problems. *Bull Amer Math Soc (NS)* 22(1):85–90
- de la Llave R, González A, Jorba À, Villanueva J (2005) KAM theory without action-angle variables. *Nonlinearity* 18(2):855–895
- Dumas HS (2014) The KAM story. A friendly introduction to the content, history, and significance of classical Kolmogorov – Arnold–Moser theory. World Scientific, Hackensack, xvi+361 pp. ISBN: 978-981-4556-58-3
- Gallavotti G (1983) *The elements of mechanics*. Springer, New York
- Giorgilli A, Locatelli U (1997) On classical series expansion for quasi-periodic motions. *MPEJ* 3(5):1–25
- Grébert B, Kappeler T (2014) The defocusing NLS equation and its normal form. *European Mathematical Society (EMS), Zrich*
- Herman M (1986) Sur les courbes invariantes pour les difféomorphismes de l’anneau. *Astérisque* 2:248
- Herman M (1987) Recent results and some open questions on Siegel’s linearization theorems of germs of complex analytic diffeomorphisms of  $C^n$  near a fixed point. In: *VIIIth international congress on mathematical physics* (Marseille, 1986). World Scientific Publishing, Singapore, pp 138–184
- Herman M (1989) Inégalités a priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques. *Publications Mathématiques de l’IHÉS* 70:47–101
- Hofer H, Zehnder E (1995) Symplectic invariants and Hamiltonian dynamics, the Floer memorial volume, *Progress in mathematics*, vol 133. Birkhäuser, Basel, pp 525–544
- Kappeler T, Pöschel J (2003) *KdV & KAM*. Springer, XIII, pp 279
- Koch H (1999) A renormalization group for Hamiltonians, with applications to KAM tori. *Erg Theor Dyn Syst* 19:475–521

## Books and Reviews

- Arnold VI, Novikov SP (eds) (1985) *Dynamical systems IV, Symplectic geometry and its applications*. Series: *Encyclopaedia of mathematical sciences*, vol 4, Volume package *Dynamical Systems*, Original Russian

- Kuksin SB (1993) Nearly integrable infinite dimensional Hamiltonian systems. *Lecture Notes in Mathematics*, vol 1556. Springer, Berlin
- Lanford OE III (1984) Computer assisted proofs in analysis. *Physics A* 124:465–470
- Laskar J (1996) Large scale chaos and marginal stability in the solar system. *Celest Mech Dyn Astron* 64:115–162
- Lazutkin VF (1993) KAM theory and semiclassical approximations to eigenfunctions. Springer-Verlag, Berlin
- MacKay RS (1985) Transition to chaos for area-preserving maps. *Lect Notes Phys* 247:390–454
- Meyer KR, Schmidt D (eds) (1991) Computer aided proofs in analysis. Springer-Verlag, Berlin
- Moser JK (1973) Stable and random motions in dynamical systems. With special emphasis on celestial mechanics, Hermann Weyl Lectures, the Institute for Advanced Study, Princeton, N. J. *Annals of Mathematics Studies*, no. 77. Princeton University Press/University of Tokyo Press, Princeton/Tokyo
- Pöschel J (2001) A lecture on the classical KAM theorem, Smooth ergodic theory and its applications (Seattle, WA, 1999). *Proceedings of symposia in pure mathematics*, vol 69. American Mathematical Society, Providence, pp 707–732
- Salamon D, Zehnder E (1989) KAM theory in configuration space, comment. *Math Helvetici* 64:84–132
- Sevryuk MB (1995) KAM-stable Hamiltonians. *J Dyn Control Syst* 1(3):351–366
- Sevryuk MB (1999) The lack-of-parameters problem in the KAM theory revisited, Hamiltonian systems with three or more degrees of freedom (SÁgaró, 1995), NATO advanced science institutes series C: mathematical and physical sciences, vol 533. Kluwer Academic Publishers, Dordrecht, pp 568–572
- Siegel CL, Moser JK (1971) Lectures on celestial mechanics. Springer-Verlag, Berlin
- Weinstein A (1979) Lectures on symplectic manifolds, 2nd ed. In: CBMS regional conference series in mathematics, vol 29. American Mathematical Society, Providence
- Xu J, You J, Qiu Q (1997) Invariant tori for nearly integrable Hamiltonian systems with degeneracy. *Math Z* 226(3):375–387
- Yoccoz JC (1992) Travaux de Herman sur les tores invariants, Séminaire Bourbaki, (1991/92) Exp. 754. *Astérisque* 206:311–344
- Yoccoz J-C (1995) Théorème de Siegel, pôlynomes quadratiques et nombres de Brjuno. *Astérisque* 231:3–88



## Nekhoroshev Theory

Laurent Niederman

Topologie et Dynamique – UMR 8628 du CNRS,  
Université Paris, Paris, France

Astronomie et Systèmes Dynamiques – UMR  
8028 du CNRS, IMCCE, Paris, France

### Article Outline

Glossary

Definition of the Subject

Introduction

Integrable and Quasi-Integrable Hamiltonian  
Systems

Averaging Principle

Hamiltonian Perturbation Theory

Exponential Stability of Constant Frequency  
Systems

Nekhoroshev Theory (Global Stability)

Application of the Global Nekhoroshev Theory

Future Directions

Appendix A

Bibliography

### Glossary

**Quasi integrable Hamiltonian** A Hamiltonian is quasi integrable if it is close to another Hamiltonian whose associated system is integrable by quadrature. Here, we will consider real analytic Hamiltonians on a domain  $\mathcal{D}$  which admit a holomorphic extension on a complex strip  $\mathcal{D}_{\mathbb{C}}$  around  $\mathcal{D}$  and the closeness to the integrable Hamiltonian is measured with the supremum norm  $\|\cdot\|_{\infty}$  over  $\mathcal{D}_{\mathbb{C}}$ .

**Exponentially stable Hamiltonian** An integrable Hamiltonian governs a system which admits a collection of first integral. We say that an integrable Hamiltonian  $h$  is exponentially stable if for any small enough Hamiltonian perturbation of  $h$ , the solutions of the

perturbed system are at least defined over time-scales which are exponentially long with respect to the inverse of the size of the perturbation. Moreover, the first integrals of the integrable system should remain nearly constant along the solutions of the perturbed system over the same amount of time.

**Nekhoroshev theorem (1977)** There exists a *generic set* of real analytic integrable Hamiltonians which are *exponential stable*.

### Definition of the Subject

We only know explicitly the solutions of the Hamiltonian systems which are integrable by quadrature. Unfortunately these integrable Hamiltonians are exceptional but many physical problems can be studied by a Hamiltonian system which differs from an integrable one by a small perturbation. One of the most famous is the motions of the planets around the Sun which can be seen as a perturbation of the integrable system associated to the motion of noninteracting points around a fixed attracting center. Poincaré (Ramis and Schäfke 1996) considered the study of these quasi integrable Hamiltonian systems as the *Problème général de la dynamique*. This question was tackled with the Hamiltonian perturbation theories which were introduced at the end of the nineteenth century precisely to study the planetary motions (see “Hamiltonian Perturbation Theory (and Transition to Chaos)”).

Significant theorems concerning the mathematical justification of these methods were not put forward until the 1950s and later. A cornerstone of these results is the Kolmogorov–Arnold–Moser (KAM) theory which states that under suitable nondegeneracy and smoothness assumptions, for a small enough perturbation most of the solutions of a nearly integrable Hamiltonian system are quasi periodic, hence they are stable over infinite times. More accurately, KAM theory provides a set of large measure of invariant tori which support stable and global solutions of the perturbed system. But

this set of stable quasi periodic solutions has a complicated topology: it is a Cantor set nowhere dense and without interior points. Numerically, it is extremely difficult to determine if a given solution is quasi periodic or not.

Moreover, for a  $n$  degrees of freedom Hamiltonian system (hence a  $2n$ -dimensional system) KAM theory provides  $n$ -dimensional tori. If  $n = 2$ , these two-dimensional invariant tori divide the three-dimensional energy level, therefore the solutions of the perturbed system are global and bounded over *infinite* times but an arbitrary large drift of the orbits is still possible for  $n \geq 3$ . Actually, Arnold (1964) proposed examples of quasi integrable Hamiltonian systems where an arbitrary large instability occurs for an arbitrary small perturbation. This is known as “Arnold diffusion”.

Thus, results of stability for quasi integrable Hamiltonian systems which are valid for an open set of initial condition can only be proved over *finite* times.

The first theorems of effective stability over finite but very long times were proved by Moser (Neishtadt 1981) and Littlewood (Lochak 1992) around an elliptic equilibrium point in a Hamiltonian system. Extending this kind of result in a general framework, Nekhoroshev (Niederman 1996, 1998) proved in 1977 a fundamental theorem of global stability which completes KAM theory and states that, for a *generic set* of integrable Hamiltonian, Arnold diffusion can only occur over exponentially long times with respect to the inverse of the size of the perturbation.

These theories have been applied extensively in celestial mechanics but also to study molecular dynamics, beam dynamics, billiards, geometric-numerical integrators, stability of certain solutions of nonlinear PDE (Schrodinger, wave, ...).

## Introduction

We first recall that the canonical equations with a given scalar function  $\mathcal{H}$  (the Hamiltonian) which is differentiable on an open set  $\mathcal{D} \subset \mathbb{R}^{2n}$  equipped with the symplectic form of Liouville  $\omega = \sum_i dp_i \wedge dq_i$  for  $(q_1, \dots, q_n) \in \mathbb{R}^n$  and their conjugate coordinates  $(p_1, \dots, p_n) \in \mathbb{R}^n$  can be written:

$$\dot{p} = -\partial_q \mathcal{H}(p, q); \quad \dot{q} = \partial_p \mathcal{H}(p, q).$$

Actually, this kind of system can be defined over any symplectic manifold.

Moreover, a diffeomorphism  $F$  is symplectic or *canonical* if it preserves Liouville form:  $\Phi^* \omega = \omega$ . Especially, for a Hamiltonian  $X$ , the flow  $\Phi_X^t$  linked to the canonical system governed by  $X$  is a symplectic transformation over its domain of definition (Moan 2004). Such a diffeomorphism preserves the Hamilton equations: in the new variables  $(p, q) = \Phi(P, Q)$  the transformed system is still canonical with the initial Hamiltonian  $K(P, Q) = \mathcal{H} \circ \Phi(P, Q)$ .

## Integrable and Quasi-Integrable Hamiltonian Systems

According to the theorem of Liouville–Arnold, under general topological conditions, a Hamiltonian system integrable by quadrature can be reduced to a system defined over  $\Omega \times \mathbb{T}^n$  for an open set  $\Omega \subset \mathbb{R}^n$  and the  $n$ -dimensional torus  $\mathbb{T}^n$  with a Hamiltonian which does not depend on the  $n$  angles. The new variables  $(I, \theta) \in \Omega \times \mathbb{T}^n$  are called actions-angle variables and the system is linked to a Hamiltonian  $K(I)$ , hence the equations take the trivial form  $\dot{I} = 0, \dot{\theta} = \nabla K(I)$  where  $\nabla K$  is the gradient of  $K$  and we obtain quasi-periodic motions of frequencies  $\nabla K(I)$ .

If a small Hamiltonian perturbation is added to an integrable system, then in action-angle variables the considered system is governed by

$$\mathcal{H}(\varepsilon, I, \theta) = h(I) + \varepsilon f(I, \theta) \quad (1)$$

where  $\mathcal{H} \in C^\omega(\Omega \times \mathbb{T}^n, \mathbb{R})$

and the equations become:

$$\begin{aligned} \dot{I} &= -\varepsilon \partial_\theta f(I, \theta); \quad \dot{\theta} \\ &= \nabla h(I) + \varepsilon \partial_I f(I, \theta). \end{aligned} \quad (2)$$

Hence, the variables are separated into two groups: the fast variables and the other variables which evolve slowly (the actions but also the angles with zero frequencies), over times of order  $1/\varepsilon$  their evolution may be of order 1.

## Averaging Principle

The *averaging principle* consists of the replacement of the initial system by its time average along the unperturbed flow  $\Phi_h^t$  linked to  $h$  which means that we consider the averaged Hamiltonian:

$$\langle \mathcal{H} \rangle(I, \theta) = h(I) + \varepsilon \langle f \rangle(I, \theta)$$

$$\text{with } \langle f \rangle(I, \theta) = \lim_{t \rightarrow \infty} \left( \frac{1}{t} \int_0^t f(I, \theta + s \nabla h(I)) ds \right).$$

Actually, this average depends on the commensurability relations which are satisfied by the components of vector  $\nabla h(I)$ .

More accurately, to a submodule  $\mathcal{M} \subset \mathbb{Z}^n$ , we associate its resonant zone:

$$\begin{aligned} Z_{\mathcal{M}} &= \{I \in \mathbb{R}^n \text{ such that } k \cdot \nabla h(I) = 0 \\ &\text{if and only if } k \in \mathcal{M}\}. \end{aligned} \quad (3)$$

Let  $r \in \{0, \dots, n-1\}$  be the rank of  $\mathcal{M}$ , there exists a unimodular matrix  $\mathcal{R} \in \text{SL}(n, \mathbb{Z})$  such that  $\omega \in \mathcal{M}^\perp$  if and only if the first  $r$  lines of  $\mathcal{R}\omega$  are zeros. Hence, the symplectic transformation  $\phi = \mathcal{R}\theta$  and  $I = {}^t\mathcal{R}^{-1}J$  defined over  $\Omega \times \mathbb{T}^n$  yields the new Hamiltonian  $\tilde{h}(J) + \varepsilon \tilde{f}(J, \phi)$  where  $J = (J_1, J_2) \in \mathbb{R}^r \times \mathbb{R}^{n-r}$  and  $\phi = (\phi_1, \phi_2) \in \mathbb{T}^r \times \mathbb{T}^{n-r}$  such that  $\nabla \tilde{h}(J) = (0, \omega(J))$  when  ${}^t\mathcal{R}^{-1}J \in Z_{\mathcal{M}}$ . Moreover, the second component  $\omega(J) \in \mathbb{R}^{n-r}$  does not satisfy any commensurability relation.

Equivalently, on the set  $Z_{\mathcal{M}}$ , we can consider the torus  $\mathbb{T}^n$  as the product  $\mathbb{T}^r \times \mathbb{T}^{n-r}$  where the unperturbed flow is *constant* over  $\mathbb{T}^r$  and *ergodic* over  $\mathbb{T}^{n-r}$ .

Hence, at infinity, the time average  $\langle \tilde{f} \rangle(J, \phi)$  is equal to the space average:

$$\langle \tilde{f} \rangle(J_1, J_2, \phi_1) = \left( \frac{1}{2\pi} \right)^{n-r} \int_{\mathbb{T}^{n-r}} \tilde{f}(J_1, J_2, \phi_1, \phi_2) d\phi_2, \quad (4)$$

and the averaged system involves only variables which evolve slowly. For instance, to determine the solutions numerically, we only need to take a step of integration of order  $1/\varepsilon$  (Ilyashenko 1986).

Actually, in the expression (4), we have only made a *partial averaging* by removing  $n - r$  angles. Hence, the considered system is simpler but, for a generic perturbation, it is *only* for the zero module  $\mathcal{M} = \{0\}$  that we obtain an integrable Hamiltonian.

It should be noticed that, for an arbitrary module  $\mathcal{M}$ , the solutions of the average system can be unbounded. For instance, the Hamiltonian  $\mathcal{H}(I_1, I_2, \theta_1, \theta_2) = I_1 I_2 + \varepsilon \sin(\theta_2)$  is averaged with respect to the module  $\mathcal{M} = \mathbb{Z} \vec{\mathbf{i}}$  with  $\vec{\mathbf{i}} = (1, 0)$  since  $\theta_1$  does not appear in the perturbation and it admits the unbounded solution  $(I_1(t), I_2(t), \theta_1(t), \theta_2(t)) = (0, -\varepsilon t, \varepsilon t^2/2, 0)$  which starts from the origin  $(0, 0, 0, 0)$  at  $t = 0$ . One can also notice that we have a maximal speed of drift of the action  $I_2$  with respect to the size of the perturbation.

On the other hand, a key observation for the proof of Nekhoroshev theorem is the fact that for an arbitrary module  $\mathcal{M}$ , the canonical equations ensure that the variations of the actions under the averaged vector field with respect to  $\mathcal{M}$  are located in the *subspace spanned by*  $\mathcal{M}$ . This point will be specified in section “The Initial Statement”, it will be the only place in Nekhoroshev’s proof where the canonical form of the equations is used.

## Hamiltonian Perturbation Theory

The averaging principle is based on the idea that the oscillating terms discarded in averaging cause only small oscillations which are superimposed to the solutions of the averaged system. In order to prove the validity of the averaging principle, one should check that any solutions of the perturbed system remain close to the solution of the averaged system with the same initial condition. Especially, this will be the case if one finds a canonical transformation  $\varepsilon$ -close to identity which transforms the perturbed Hamiltonian to its average. Hence we are reduced to a problem of normal form where one looks for a convenient system of coordinates which gives the simplest possible form to the considered system.

Here, we consider the transformation  $\Phi_X^1$  given by the time 1 flow of the Hamiltonian system governed by  $X(I, \theta) = \varepsilon X_1(I, \theta)$ .

With Taylor formula, the transformed Hamiltonian  $\mathcal{H} \circ \Phi_X^1$  admits the following expansion with respect to  $\varepsilon$ :

$$\mathcal{H} \circ \Phi_X^1 = h + \varepsilon(f + \nabla h(I) \cdot \partial_\theta X_1(I, \theta)) + O(\varepsilon^2),$$

and in order to obtain  $\mathcal{H} \circ \Phi_X^1 = h(I) + \varepsilon\langle f \rangle$  one must solve:

$$\nabla h(I) \cdot \partial_\theta X_1(I, \theta) = -f + \langle f \rangle, \quad (5)$$

which is the homological equation, this is the central equation of perturbation theory.

Since we are in the analytic setting, the function  $f$  admits the expansion

$$\sum_{k \in \mathbb{Z}^n} f_k(I) \exp(ik\theta).$$

Hence, for an averaging with respect to a resonant module  $\mathcal{M}$  around the resonant zone  $\mathcal{Z}_{\mathcal{M}}$  linked to  $\mathcal{M}$ , the homological equation admits the formal solution:

$$X_1(I, \theta) = \sum_{k \notin \mathcal{M}} \frac{f_k(I)}{i(k \cdot \nabla h(I))} \exp(ik\theta), \quad (6)$$

thus one obtains a transformation which normalizes the Hamiltonian at first order in  $\varepsilon$ .

In the same way, one can *formally* eliminate the fast angles at all orders by looking at a transformation generated by a Hamiltonian  $X(I, \theta) = \sum_{n \geq 1} \varepsilon^n X_n(I, \theta)$ . Indeed, the same type of homological equation appears at all order to determine  $X_n$  for  $n > 1$ .

This is the *Linstedt method*.

With the previous construction, it is obvious that the normalizing transformation admits an expansion which is divergent if the denominators  $k \cdot \nabla h(I)$  for  $k \notin \mathcal{M}$  are too close to zero on the considered domain in the action space. This is the well-known problem of the small denominators which was emphasized by Poincaré (Ramis and Schäfke 1996) in his celebrated theorem about nonexistence of analytic first integrals for a generic quasi integrable Hamiltonian.

It is less known that even without small denominators, the classical perturbation theory can yield divergent expansions. This is the problem of the great multipliers according to Poincaré terminology which come from the successive differentiations.

This problem is presented in the next subsection.

## Exponential Stability of Constant Frequency Systems

### The Case of a Single Frequency System

The normalization of an analytic quasi-integrable Hamiltonian system with only *one fast* phase is one of the main problems in perturbation theory. This question appears naturally to compute the time of approximate conservation of the adiabatic invariants (Bourgain 2004; Lochak and Neishtadt 1992). It has been accurately studied in (Nekhoroshev 1977; Valdinoci 2000) and we will focus our attention on this case where the phenomenon of divergence without small denominators appears in its simplest setting.

Indeed, as in the section “Hamiltonian Perturbation Theory”, one can build *formally* the Hamiltonian  $X(I, \theta) = \sum_{n \geq 1} \varepsilon^n X_n(I, \theta)$  which generates a normalizing symplectic transformation and eliminates the fast angle in the perturbation. But, for a generic analytic quasi-integrable Hamiltonian, it can be shown (Nekhoroshev 1977; Valdinoci 2000) that perturbation theory yields a Gevrey-2 normalizing transformation over  $U \times \mathbb{T}^n$  (i.e.:  $\mathcal{T} \in C^\infty(U \times \mathbb{T}^n)$  with  $\|\partial^k \mathcal{T}\|_\infty \leq CM^{2|k|} k!^2$  where  $C, M$  are positive constants and  $k = (k_1, \dots, k_{2n}) \in \mathbb{N}^{2n}$ ;  $|k| = |k_1| + \dots + |k_{2n}|$ ;  $k! = k_1! \dots k_{2n}!$ ) such that the initial perturbed Hamiltonian  $\mathcal{H} = h + \varepsilon f$  is transformed into an integrable Hamiltonian  $h_\varepsilon(\hat{I})$  with a one parameter family  $h_\varepsilon$  of scalar functions Gevrey-2 over  $U$ .

Hence, the normalizing transformations are usually divergent. On the other hand, by general properties of Gevrey functions (see section 3.3 in Meyer and Hall 1992, and the references therein) if one considers the transformation generated by the truncated expansion

$$\sum_{n \geq 1}^N \varepsilon^n X_n(I, \theta)$$

obtained after  $N$  steps of perturbation theory, then the transformed Hamiltonian is normalized up to a remainder of size  $\varepsilon^{N+1}N!$ .

In the appendix, we will study an example of a quasi-integrable Hamiltonian where this latter estimate cannot be improved and where the source of divergence of the normalizing transformation which comes from the successive differentiations in the construction can be emphasized.

We see that the remainder of size  $\varepsilon^{N+1}N!$  decreases rapidly before increasing to infinity, following Poincaré (Ramis and Schäfke 1996) this is a “convergent expansion according to astronomers” and a “divergent expansion according to geometers”.

Now, we can use the process of “summation at the smallest term”: for a fixed  $\varepsilon > 0$ , one obtains an optimal normalization with a truncation at order  $N$  such that

$$\|\partial_{\theta_2} X_{N-1}\|_{\infty} \simeq \varepsilon \|\partial_{\theta_2} X_N\|_{\infty}$$

which yields  $N = E(1/\varepsilon)$ .

Finally, the Stirling formula yields the size of the remainder:  $\sqrt{2\pi\varepsilon} \exp(-1/\varepsilon)$  which is *exponentially small* with respect to the inverse of the size of the perturbation.

More generally, Marco and Sauzin (Meyer and Hall 1992) have proved in the same setting that starting from a Gevrey- $a$  Hamiltonian ( $\|\partial^k \mathcal{H}(I, \theta)\|_{\infty} \leq CM^{\alpha+|k|}(k!)^{\alpha}$ ), one can build a normalizing transformation which is Gevrey- $a+1$  and these estimates cannot be improved usually but the previous construction is still possible. Indeed, one can still make a summation at the smallest term and obtain a canonical change of coordinates which normalizes the Hamiltonian up to an exponentially small remainder with respect to the size of the perturbation.

In the case where the averaged Hamiltonian is *integrable*, according to the mean value theorem, the speed of drift of the normalized action variables is at most *exponentially slow*. Since the size of the normalizing transformation is of order  $\varepsilon$ , the

initial actions admit at most a drift of size  $\varepsilon$  over an *exponentially long time*.

### The Case of a Strongly Nonresonant Constant Frequency System

Systems with constant frequencies, hence  $h(I) = \omega \cdot I$  for some constant vector  $\omega \in \mathbb{R}^n$ , appear when we consider small nonlinear interactions of linear oscillatory systems or the action of quasi-periodic perturbations on linear oscillatory systems. In any case, the considered Hamiltonian can be written:  $\mathcal{H} = \omega \cdot I + f(I, \theta)$  with a small function  $f \in C^{\omega}(\Omega \times \mathbb{T}^n, \mathbb{R})$  where  $\Omega$  is an open set in  $\mathbb{R}^n$ .

Moreover, we assume here that the frequency? is a  $(\gamma, \tau)$ -Diophantine vector for some positive constants? and  $\tau$ , hence  $\omega \in \Omega_{\gamma, \tau}$ :

$$\Omega_{\gamma, \tau} = \left\{ \omega \in \mathbb{R}^n \text{ such that } |k \cdot \omega| \geq \frac{\gamma}{\|k\|_{\infty}^{\tau}} \right. \quad (7) \\ \left. \text{for all } k \in \mathbb{Z}^n \setminus \{(0, \dots, 0)\} \right\}.$$

We recall that the measure of the complementary set of  $\Omega_{\gamma, \tau}$  is of order  $O(\gamma)$  for  $\tau > n-1$ .

Under these assumptions, one can prove (Benettin and Giorgilli 1994; Giorgilli and Morbidelli 1997; Lochak 1992; Steichen and Giorgilli 1998) that for a small enough analytic perturbation, the action variables of the unperturbed problem become quasi integrals of the perturbed system over exponentially long times, more specifically:

**Theorem 1** Consider a Hamiltonian  $\omega \cdot I + f(I, \theta)$  real analytic over a domain  $\mathcal{U} \times \mathbb{T}^n \subset \mathbb{R}^n \times \mathbb{T}^n$  which admit a holomorphic extension on a complex strip of width  $\rho > 0$  around  $\mathcal{U} \times \mathbb{T}^n$  in  $\mathbb{C}^{2n}$  (Benettin and Giorgilli 1994; Giorgilli and Morbidelli 1997; Lochak 1992; Steichen and Giorgilli 1998).

The supremum norm for a holomorphic function on this complex strip is denoted  $\|\cdot\|_{\rho}$ .

There exists positive constants  $C_1, C_2, C_3, C_4$  which depend only on  $\gamma, \tau, \rho, n$  such that if  $\varepsilon = \|f\|_{\rho} < C_1\gamma$ , an arbitrary solution  $(I(t), \varphi(t))$  of the perturbed system associated to  $\omega \cdot I + f(I, \theta)$  with an initial action  $I(t_0) \in \mathcal{U}$  is defined at least over an exponentially long time and satisfies:

$$\begin{aligned} \|I(t) - I(0)\| &\leq C_2 \varepsilon \text{ if } |t| \\ &\leq C_3 \exp\left(C_4 \varepsilon^{-\frac{1}{1+\varepsilon}}\right). \end{aligned} \quad (8)$$

The proof is based on the existence of a normalizing transformation up to an exponentially small error. This is possible since we have lower bounds on the small denominators and the growth of the coefficients in the normalizing expansion is reduced to a combinatorial problem.

Finally, since the averaged Hamiltonian is integrable, the speed of drift of the action variables is at most exponentially slow.

## Nekhoroshev Theory (Global Stability)

### The Initial Statement

Thirty years ago, Nekhoroshev (Niederman 1996, 1998) stated a *global* result of stability which is valid for a *generic* set of integrable Hamiltonian. Especially, we don't have anymore a control on the small denominators as in the previous section but we have to handle the resonant zones. Nekhoroshev's reasonings allow one to prove a global result of stability *independent* of the arithmetical properties of the unperturbed frequencies by taking into account the *geometry of the integrable system*. This is really a change of perspective with respect to the previous results. The key ingredient is to find a suitable property of the integrable Hamiltonian, namely the property of *steepness* introduced in the sequel, which ensures that a drift of the actions in the averaged system with respect to a module  $\mathcal{M} \subset \mathbb{Z}^n$  leads to an *escape* of the resonant zone  $\mathcal{Z}_{\mathcal{M}}$ .

More specifically, Nekhoroshev proved global results of stability over open sets of the following type:

**Definition 2** (exponential stability) Consider an open set  $\Omega \subset \mathbb{R}^n$ , an analytic integrable Hamiltonian  $h : \Omega \rightarrow \mathbb{R}$  and action-angle variables  $(I, \varphi) \in \Omega \times \mathbb{T}^n$  where  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ .

For an arbitrary  $\rho > 0$ , let  $O_\rho$  be the space of analytic functions over a complex neighborhood  $\Omega_\rho \subset \mathbb{C}^{2n}$  of size  $\rho$  around  $\Omega \times \mathbb{T}^n$  equipped with the supremum norm  $\|\cdot\|_\rho$  over  $\Omega_\rho$ .

We say that the Hamiltonian  $h$  is exponentially stable over an open set  $\tilde{\Omega} \subset \Omega$  if there exists positive constants  $\rho$ ,  $C_1$ ,  $C_2$ ,  $a$ ,  $b$  and  $\varepsilon_0$  which depend only on  $h$  and  $\tilde{\Omega}$  such that:

- i)  $h \in O_\rho$ .
- ii) For any function  $\mathcal{H}(I, \varphi) \in O_\rho$  such that  $\|\mathcal{H} - h\|_\rho = \varepsilon < \varepsilon_0$ , an arbitrary solution  $(I(t), \varphi(t))$  of the Hamiltonian system associated to  $\mathcal{H}$  with an initial action  $I(t_0)$  in  $\tilde{\Omega}$  is defined over a time  $\exp(C_2/\varepsilon^a)$  and satisfies:

$$\begin{aligned} \|I(t) - I(t_0)\| &\leq C_1 \varepsilon^b \text{ for } |t - t_0| \\ &\leq \exp(C_2/\varepsilon^a), \end{aligned} \quad (9)$$

$a$  and  $b$  are called stability exponents.

**Remark** Along the same lines, the previous definition can be extended to an integrable Hamiltonian in the Gevrey class (see Meyer and Hall 1992).

Hence, for a small enough perturbation, the action variables of the unperturbed problem become *quasi integrals* of the perturbed system over exponentially long times.

In order to introduce the problem, we begin by a typical example of *non-exponentially* stable integrable Hamiltonian:  $h(I_1, I_2) = I_1 I_2$ . Indeed, the perturbed system governed by  $h(I_1, I_2) + \varepsilon \sin(\theta_2)$  admits the unbounded solution  $(I_1(t), I_2(t), \theta_1(t), \theta_2(t)) = (0, \varepsilon t, \varepsilon t^2/2, 0)$  which starts from the origin  $(0, 0, 0, 0)$  at  $t = 0$ , hence a drift of the actions  $(I_1(t), I_2(t))$  on a segment of length 1 occurs over a timespan of order  $1/\varepsilon$ .

The important feature in this example which has to be avoided in order to ensure exponential stability is the fact that the gradient  $\nabla h(I_1, 0)$  remains orthogonal to the first axis. Equivalently, the gradient of the restriction of  $h$  on this first axis is identically zero.

Nekhoroshev (Nekhoroshev 1979; Niederman 1996, 1998) introduced the class of steep functions where this problem is avoided. The property of steepness is a quantitative condition of transversality for a real valued function differentiable



over an open set  $\Omega \subset \mathbb{R}^n$  which involves *all* the affine subspaces which intersect  $\Omega$ . Actually, steepness can be characterized by the following simple geometric criterion proved thanks to theorems of real subanalytic geometry (Poincaré 1892):

**Theorem 3** *A real analytic scalar function without critical points is steep if and only if its restriction to any proper affine subspace admits only isolated critical points (Poincaré 1892).*

This is an extension of a previous similar result in the holomorphic case (Kuksin and Pöschel 1994).

In this setting, Nekhoroshev proved the following:

**Theorem 4** *If the integrable Hamiltonian  $h$  is real analytic, does not admit critical points, is non-degenerate ( $|\nabla^2 h(I)| \neq 0$  for any  $I \in \Omega$ ) and steep then  $h$  is exponentially stable (Niederman 1996, 1998).*

The set of steep functions is *generic* among sufficiently smooth functions. For instance, we have seen that the function  $xy$  is not steep but it can easily be shown that  $xy + x^3$  is steep. Actually, a given function can be transformed into a steep function by adding higher order terms (Nekhoroshev 1979; Niederman 1996). It can be noticed that the (quasi-)convex functions are the *steepest* functions since their restrictions to any affine subspaces admit at most *one* critical point which is also nondegenerate.

The original proof of Nekhoroshev is *global*. It is based on a covering of the action space in open sets with controlled resonance properties where one can build resonant normal forms (i.e.: where only resonant harmonics are retained) up to an exponentially small remainder. The averaged Hamiltonian is not necessarily integrable but, thanks to the steepness of the integrable Hamiltonian, if a drift of the normalized actions occurs then it can only lead to a zone associated to resonances of *lower multiplicity* than the initial one (i.e.: the resonant module admits a lower dimension). Eventually, after a short distance the orbits

reach a resonance-free area (i.e.: the Fourier expansion of the normalized perturbation admits only nonresonant harmonics up to an exponentially small remainder). Then, the local normal form is integrable and yields the confinement of the action variables over the desired amount of time.

### Improved Versions of Nekhoroshev Theorem

The articles of Nekhoroshev remained largely unnoticed in the western countries until Benettin, Galgani and Giorgilli (Benettin and Gallavotti 1986) rewrote and clarified the initial proof in the convex case.

Benettin and Gallavotti (Benettin and Giorgilli 1994) proved that under an assumption of (quasi) convexity of the unperturbed Hamiltonian, the proofs of these theorems can be simplified. Indeed, after an averaging as in the steep case, the quasi convexity and the *energy conservation* ensure that the normalized Hamiltonian is an *approximate Liapunov function* over exponentially long time intervals. This allows one to confine the actions in the initial set where the considered orbit was located (we do not have to consider a drift over resonant areas of different multiplicity as in the original proof). Hence, the construction of a *single* normal form is enough to confine the actions in the convex case.

Following this idea, Lochak (1993, 1995) has significantly simplified the proof of Nekhoroshev estimates for the convex quasi integrable Hamiltonians. His reasonings are based on normalization around the periodic orbits of the integrable Hamiltonian which represent the worst type of resonances. Using convexity, Lochak obtains open sets around the periodic orbits which hold exponential stability. Then, Dirichlet theorem about *simultaneous Diophantine approximation* ensures that these open sets recover the whole action space and yield the global result.

A remarkable feature of this proof is the fact that *improved estimates* can be obtained in the vicinity of resonances thanks to the relative abundance of periodic orbits in these areas. More specifically, periodicity corresponds to  $n - 1$  commensurability relations and we have already several commensurability relations at the



resonances hence Dirichlet theorem can be applied on a lower dimensional space with better rates of approximation (Lochak 1993, 1995). These improvements are important to extend Nekhoroshev estimates for large systems or infinite dimensional systems, they also fit with the speed of drift of the action variables in examples of unstable quasi integrable Hamiltonian (these points will be discussed in the sequel).

It can also be noticed that averaging along the periodic orbits of the integrable system is exactly a *one phase* averaging without small denominators.

Toward sharp estimates, Lochak–Neishtadt (Lochak et al. 1994) and Pöschel (Steichen and Giorgilli 1998) have independently obtained the following:

**Theorem 5** If the integrable Hamiltonian  $h$  is real analytic, does not admit critical points and convex over a domain  $\Omega \subset \mathbb{R}^n$  then  $h$  is exponentially stable over  $\Omega$  with the global exponents  $a = b = 1/2n$  (Lochak et al. 1994; Steichen and Giorgilli 1998).

Moreover, around the resonant zones linked to a module of rank  $m < n$ , the integrable Hamiltonian  $h$  is exponentially stable with the improved exponents  $a = b = 1/2(n - m)$ .

The proof in (Lochak et al. 1994), explicitly derived in (Marco and Lochak 2005), relies on Lochak periodic orbits method together with a refined procedure of averaging due to Neishtadt (Nekhoroshev 1977). In (Steichen and Giorgilli 1998), the original scheme of Nekhoroshev is combined with a refined study of the geometry of resonances which gives an accurate partition of the action space in open sets where the action variables are confined and also Neishtadt's averaging procedure is used. Pöschel's study of the geometry of resonances should also be important in the study of Arnold diffusion.

This value of the time exponent ( $a = 1/2n$ ) is expected to be optimal in the convex case according to heuristic reasonings of Chirikov (Delshams and Gutierrez 1996), see also (Lochak 1993) on the speed of drift of Arnold diffusion. Actually, Marco–Sauzin (Meyer and Hall 1992) in the Gevrey cases and Marco–

Lochak (Marco and Sauzin 2003) in the analytic case have essentially proved the optimality of the improved exponent  $a = 1/2(n - 2)$  in the doubly resonant zones starting from an example of unstable quasi integrable Hamiltonian given by Michel Herman.

On the other hand, the previous studies except the original one of Nekhoroshev do not cover the cases of a time-dependent perturbation or a perturbed steep integrable Hamiltonian, despite their importance in physics. For instance, a time periodic perturbation of a convex Hamiltonian can be reduced to the time-independent perturbation of a quasi convex Hamiltonian in the extended phase space, but this is not the case for a general time-dependent perturbation, where energy conservation cannot be used. This problem has been studied in the light of Nekhoroshev theory by Giorgilli and Zehnder (Gramchev and Popov 1995) in connection with the dynamics of a particle in a time-dependent potential with a high kinetic energy.

A new general study of the stability of steep integrable Hamiltonians has been carried out in (Niederman 2007). This proof of stability relies on the mechanism of Nekhoroshev since we analyze the dynamics around resonances of any multiplicity and use local resonant normal forms but the original *global* construction is substituted with a *local* construction along each trajectory of the perturbed system. This construction is based on the approximation of the frequencies  $\nabla h(I(t))$  at certain times by rational vectors thanks to Dirichlet theorem of simultaneous Diophantine approximation as in Lochak's proof of exponential stability in the convex case. This allows significant simplifications with respect to Nekhoroshev's original proof. Moreover, the results of Lochak and Pöschel are generalized for the steep case since the exponents of stability derived in (Niederman 2007) give back the exponents  $a = b = 1/2n$  in the particular case of convex integrable Hamiltonians.

In (Pöschel 1993), Nekhoroshev estimates are proved for perturbations of integrable Hamiltonians which satisfy a *strictly weaker* condition of nondegeneracy than the initial condition of steepness. Indeed, the lack of steepness allows a drift of

the action variables around the resonant zones. But, due to the exponential decay of the Fourier coefficients in the expansion of the Hamiltonian vector field, such a drift would be extremely slow if one consider resonant zones linked to a module spanned by integer vectors of large length. Thanks to this property, Morbidelli and Guzzo (Moser 1955) have observed that the Hamiltonian  $h(I_1, I_2) = I_1^2 - \delta I_2^2$  where  $\delta$  is the square of a Diophantine number is *non steep* but nevertheless  $h$  is *exponentially stable* since its isotropic directions are the lines spanned by  $(1, \pm\sqrt{\delta})$  which are “far” from the lines with rational slopes. This phenomenon has been studied numerically for the quadratic integrable Hamiltonians in (Guzzo and Morbidelli 1997).

Starting from this observation, a general weak condition of steepness which involves only the affine subspaces spanned by *integer* vectors has been stated in (Pöschel 1993) with a complete proof of exponential stability in this setting.

The point in this refinement lies in the fact that it allows one to exhibit a *generic* class of real analytic integrable Hamiltonians which are exponentially stable with *fixed* exponents of stability  $a$  and  $b$  while Nekhoroshev original theory provides a generic set of exponentially stable integrable Hamiltonians but with exponents of stability which can be *arbitrarily small* (Pöschel 1993).

More specifically, we consider genericity in a measure theoretical sense since:

**Theorem 6** Consider an arbitrary real analytic integrable Hamiltonian  $h$  defined on a neighborhood of the closed ball  $\overline{B}_R^{(n)}$  of radius  $R$  centered at the origin in  $\mathbb{R}^n$  (Pöschel 1993).

For almost any  $\Omega \in \mathbb{R}^n$ , the integrable Hamiltonian  $h_\Omega(x) = h(I) - \Omega \cdot I$  is exponentially stable with the exponents:

$$a = \frac{b}{2 + n^2} \quad \text{and} \quad b = \frac{1}{2(1 + 2^n n)}.$$

Finally, all these results can be generalized for quasi integrable *symplectic mappings* thanks to a theorem of inclusion of an analytic symplectic diffeomorphism into the flow linked to a real

analytic Hamiltonian up to an exponentially small accuracy (Kuksin 2006; Arnold et al. 2006 or Guzzo and Benettin 2001 for a direct proof).

### KAM Stability, Exponential Stability, Nekhoroshev Stability

A last point which should be emphasized is the link between KAM stability, exponential stability and Nekhoroshev stability which are cornerstones in the study of stability of analytic quasi integrable Hamiltonian systems.

A first problem is the stability of the solutions in a neighborhood of a Lagrangian invariant torus over which an analytic Hamiltonian vectorfield induces a linear flow of frequency?. Then the considered Hamiltonian can be written as  $\mathcal{H}(I, \theta) = \omega \cdot I + \mathcal{F}(I, \theta)$  in action-angle variables where the perturbation  $\mathcal{F}(I, \theta)$  is analytic in a neighborhood of the origin and starts at order *two* in actions.

In the case of a KAM torus, the frequency is strongly nonresonant (Diophantine). With a suitable rescaling, the expansion of the Hamiltonian takes the form considered in Theorem 1 and the exponential estimates of stability (9) are valid.

In the general case (where the frequency can satisfy resonances of low order), the previous procedure cannot be applied. On the other hand, if one assumes that there are no resonances of order lower or equal to *four*:

$$\forall k \in \mathbb{Z}^n \setminus \{0\} \text{ such that } |k| = |k_1| + \dots$$

$$+ |k_n| \leq 4 \text{ then } |k \cdot \omega| \neq 0,$$

it is possible to perform a Birkhoff's normalization which reduces the studied Hamiltonian to:

$$\tilde{\mathcal{H}}(\tilde{I}, \tilde{\theta}) = \omega \cdot \tilde{I} + \tilde{Q}(\tilde{I}) + \tilde{\mathcal{F}}(\tilde{I}, \tilde{\theta}), \quad (10)$$

where  $\tilde{Q}$  is a quadratic form (the torsion) and  $\tilde{\mathcal{F}}(\tilde{I}, \tilde{\theta}) = o_3(\tilde{I})$ . At this point, one introduces the steepness condition required for the application of Nekhoroshev theory by imposing that the quadratic form  $\tilde{Q}$  is *sign definite* (a weaker condition is considered in Dumas 1993).

In a neighborhood of the considered torus, the problem is now reduced to the study of perturbations of a nonlinear *convex* integrable Hamiltonian and an exponential estimate of stability can be derived. These reasonings were stated in (Niederman 1996, 2.2) and more specifically studied in (Lochak 1993; Lochak and Meunier 1988).

A remarkable result of Morbidelli and Giorgilli (Morbidelli and Guzzo 1997) clearly shows that the two previous results which come respectively from Hamiltonian perturbation theory and Nekhoroshev's theorem are *independent* and can be *superimposed*. Indeed, in the case of a strongly nonresonant invariant torus (especially for a KAM torus) which admits *moreover* a sign definite torsion, one can state results of stability over superexponentially long times. The proof starts with a Birkhoff's normal form like (10) but with an exponentially small remainder. Then, Nekhoroshev's theorem is applied with a perturbation which is already exponentially small, hence we obtain a superexponential time of stability. More specifically, provided that the Birkhoff normal form is quasi-convex, results of stability over times of the order of  $\exp(\exp(cR^{-1/\epsilon}))$  can be ensured for the solutions with an initial condition in a ball of radius  $R$  small enough around the invariant torus.

Actually, the previous results were extended (Lochak 1993; Lochak and Meunier 1988) for an *elliptic equilibrium* point or a lower dimensional torus in a Hamiltonian system except for an annoying problem of singularity in the action-angle transformation. This problem does not allow one to prove a stability theorem for a complete neighborhood of an elliptic equilibrium point. The corresponding theorems for *all initial data* were obtained in (Fassò and Lewis 2001; Niederman 2006; Treschev 1994) where Nekhoroshev's estimates were established without action-angle variables.

Finally, the relationship between KAM and Nekhoroshev theory was considered in (Dullin and Fassò 2004; Giorgilli and Skokos 1997) where the existence of a sequence of nested domains in phase space which converge to the KAM set of invariant tori for the perturbed system and over which stability estimates are valid and growth occurs in a superexponential way was

proven. Especially, on the initial domain we recover the usual statement of Nekhoroshev.

## Applications

With the remaining space, we only give glimpses of application of the previous theorems in physics and astronomy. We mainly quote surveys in thesequel and these references do not form at all a complete list.

## The Case of a Constant Frequency Integrable System

The question of stability of a perturbed single frequency system corresponds to the problem of preservation of the adiabatic invariants which appears in numerous physical problems (Arnold et al. 2006; Nekhoroshev 1973).

Especially, for the applications in plasma physics, we can mention the beautiful survey of Northrop (Pöschel 1999). The problem of charge trapping by strong nonuniform magnetic fields (Van Allen belts, magnetic bottles) can also be tackled by means of the adiabatic invariant theory (Blaom 2001).

The same question arises in connection with the problem of energy equipartition in large systems of Fermi–Pasta–Ulam type and the conjecture of Boltzmann and Jeans (Benettin and Fassò 1996; Giorgilli et al. 1989).

The problem of existence of an approximate first integral over a long time but this time for a quasi integrable symplectic mapping appears to study the effective stability of the billiard flow near the boundary of a strictly convex domain in  $\mathbb{R}^n$  (Guzzo 2003).

For several degrees of freedom, results of stability over very long times for a quasi integrable Hamiltonian system were first obtained by Littlewood (Lochak 1992) about *triangular* Lagrangian equilibria in the three bodies problem. One can reduce this question to the study of a perturbed strongly nonresonant constant frequency system (Giorgilli and Morbidelli 1997).

For effective computations, it is much more efficient to make a numerical summation at the smallest term instead of plugging the data into an abstract theorem which usually gives poor estimates. Following this scheme, Giorgilli and

different coauthors have obtained effective stability results in celestial mechanics with *realistic* physical parameters (see Benettin and Fassò 1996; Fassò et al. 1998; Giorgilli et al. 1989; Giorgilli and Zehnder 1992; Arnold 1983).

The same situation of a perturbed strongly nonresonant constant frequency system appears in the study of stability of symplectic numerical integrators (Benettin and Sempio 1994; Ilyashenko 1986; Morbidelli and Giorgilli 1995).

In the realm of PDE's, results of stability around *finite dimensional nonresonant tori* can be proved (Bambusi 1999b; Benettin et al. 1985; Carati et al. 2002; Littlewood 1959 and *Perturbation Theory for PDEs* for surveys).

The remarkable paper of Bambusi and Grebert (2006) gives an extension of the previous results for *infinite dimensional* nonresonant tori.

## Application of the Global Nekhoroshev Theory

Here, we look for global results of stability for perturbed nonlinear integrable Hamiltonian systems.

This situation appears in celestial mechanics where the unperturbed system is often *properly degenerate*, namely the number of constants of motion exceeds the number of degrees of freedom. This is the case for the Kepler problem which yields the integrable part of the planetary  $n$ -bodies problem (i.e.: the approximation of the  $n$ -bodies problem corresponding to the motion of the planets in the solar system). A study of this system in the light of Nekhoroshev's theory was given in (Niederman 2004), suggesting a modification of the original statement of Nekhoroshev by considering the proper degeneracies of this system.

The question of stability in celestial mechanics was also considered for the asteroid belt (Moser 1955) where additional degeneracies and resonances appear (see also Hairer et al. 2006).

We have seen that Nekhoroshev's theory also allows one to study stability around an elliptic equilibrium point in a Hamiltonian system with a sign definite torsion. But the most famous example of such an equilibrium in astronomy, namely the

stability of an asteroid located at the top of an equilateral triangle with the Sun and Jupiter (the Lagrangian points L4, L5) cannot be tackled with the previous theorems in the convex case. It can be shown that with the actual masses of the Sun and Jupiter, the problem of stability of the Lagrangian points could be reduced to a study of a small perturbation of an integrable Hamiltonian of three degrees of freedom whose 3-jet satisfies a condition which implies steepness (Benettin et al. 1988). It allows one to prove a confinement over very long time intervals for asteroids located close enough to the Lagrangian points L4 or L5.

Other applications of Nekhoroshev's stability at an elliptic equilibrium point are the stability of the Riemann Ellipsoid (Giorgilli 1998), the fast rotation of a rigid body (Benettin et al. 1998).

A Nekhoroshev-like theory has also been developed for beam dynamics (Dumas 2005; Efthymiopoulos et al. 2004).

From a numerical point of view, a new spectral formulation of the Nekhoroshev theorem has been introduced (Guzzo et al. 2006). This allows one to recognize whether or not the motion in a quasi-integrable Hamiltonian system is in a Nekhoroshev regime (i.e. the action coordinates are eventually subject to an exponentially slow drift) by looking at the Fourier spectrum of the solutions.

In a geometric setting, an extension of Nekhoroshev's results for perturbations of convex, *noncommutatively* integrable Hamiltonian systems has been given in (Bogolyubov and Mitropol'skij 1958).

Finally, no general results like the Nekhoroshev theorem are known yet for large Hamiltonian systems or for Hamiltonian PDE's seen as infinite dimensional Hamiltonian systems. But a number of quasi-Nekhoroshev theorems for special systems of this type have been proved mostly by Bourgain and Bambusi (see Bambusi and Giorgilli 1993 for large systems and Bambusi 1999a; Benettin 2005; Carati et al. 2002 for PDE's).

## Future Directions

The analyticity of the studied systems is only needed for the construction of the normal forms

up to an exponentially high remainder. On the otherhand, the steepness condition is generic for Hamiltonians of *finite* but sufficiently high smoothness (Pöschel 1993). It would be natural to prove the analogous stability theorems in the case of smooth functions which would give stability over *polynomially* long times.

Another question is the extension of Lochak's mechanism of stabilization around resonances for *nonquasi-convex* integrable Hamiltonians. These improved exponents are at the basis of Nekhoroshev-type results which are obtained in largesystems (Bambusi and Giorgilli 1993) or in PDE (Bambusi 1999a; Benettin 2005; Carati et al. 2002) hence their generalization would be important.

The global stability results considered so far are obviously valid only if one takes into account the most pessimistic estimations in the wholephase space. On the other hand, we have just seen that these results can be improved locally. It would be relevant to make a study of the *average* exponent of stability. It could be a *space* average of these exponents by considering all initial conditions or, in certain examples, the *time* average of these exponents by taking into account the variations of the speed of drift of the actions under Arnold's diffusion.

The applications of Nekhoroshev's theory in Astronomy is an active field either analytically or in a numerical way (Guzzo 2004).

The relevance of Nekhoroshev's estimates for statistical mechanics and thermodynamics is an important question which is tackled with the problem of energy equipartition in the Fermi-Pasta-Ulam (FPU) model (see Carati et al. 2006; Celletti and Giorgilli 1991 and the references therein).

Finally, a partial generalization of KAM theory to PDEs has been carried out during the last 20 years by many high level mathematicians, but the theory developed up to now only allows one to show that *finite* dimensional invariant tori persist under perturbation. Thus, most of the initial data are outside invariant tori. It is therefore clear that it would be very important to understand the behavior of solutions starting outside the tori. This is also related to the problem of estimating the time of existence of the solutions

of hyperbolic PDEs incompact domains, a problem that is one of the most important open questions in relation to hyperbolic PDEs. Up to now Nekhoroshev's theorem has been generalized to PDE's only in order to deal with small amplitude solutions but the obtention of really global results of stability would be a challenge (Kuksin has announced these kinds of theorems for the KdV equation, see ► [“Perturbation Theory for PDEs”](#)).

## Appendix A

### An Example of Divergence Without Small Denominators

We would like to develop the following example of Neishtadt (Nekhoroshev 1977) where the phenomena of divergence without small denominators and summation up to an exponentially small remainder appear in its simplest setting (see also Benettin 2005 for another example).

Consider the quasi integrable system governed by the Hamiltonian

$$\mathcal{H}(I_1, I_2; \theta_1, \theta_2) = I_1 - \varepsilon [I_2 - \cos(\theta_1) f(\theta_2)]$$

defined over  $\mathbb{R}^2 \times \mathbb{T}^2$ ,

(11)

with

$$f(\theta) = \sum_{m \geq 1} \frac{\alpha_m}{m} \cos(m \theta),$$

where  $\alpha_m = e^{-\alpha m}$  for  $0 < \alpha \leq 1$ , this last choice corresponds to the exponential decay of the Fourier coefficients for a holomorphic function.

Indeed,  $f'(\theta) = \operatorname{Im}(g(-\alpha + i\theta))$  where  $g(z) = (\exp(z) - 1)^{-1}$  and the complex pole of  $f$  which is closest to the real axis is located at  $-i\alpha$ .

Conversely, all real function which admits a holomorphic extension over a complex strip of width  $\alpha$  (i.e.:  $|\operatorname{Im}(z)| \leq \alpha$ ) has Fourier coefficients bounded by  $(C\alpha_m)_{m \in \mathbb{N}^*}$  for some constant  $C > 0$ .

As in the section “Hamiltonian Perturbation Theory”, it is possible to eliminate formally completely the fast angle  $\theta_1$  in the perturbation *without the occurrence* of any small denominators.

Indeed, one can consider a normalizing transformation generated by  $X(\theta_1, \theta_2) = \sum_{n \geq 1} \varepsilon^n X_n(\theta_1, \theta_2)$  where the functions  $X_n$  satisfy the homological equations:

$$\begin{aligned}\partial_{\theta_1} X_1(\theta_1, \theta_2) &= \cos(\theta_1) f(\theta_2) \text{ and} \\ \partial_{\theta_1} X_n(\theta_1, \theta_2) &= \partial_{\theta_2} X_{n-1}(\theta_1, \theta_2).\end{aligned}$$

The solutions can be written  $X_n(\theta_1, \theta_2) = \cos(\theta_1 - n\frac{\pi}{2}) f^{(n-1)}(\theta_2)$ , hence:

$$\begin{aligned}X(\theta_1, \theta_2) &= \sum_{n \in \mathbb{N}^*} \varepsilon^n \sum_{m \in \mathbb{N}^*} \frac{\alpha_m}{m^2} m^n \\ &\quad \cdot \cos\left(\theta_1 - n\frac{\pi}{2}\right) \sin\left(m\theta_2 + n\frac{\pi}{2}\right)\end{aligned}$$

and, for instance,

$$X\left(\frac{\pi}{2}, 0\right) = \sum_{m \in \mathbb{N}^*} \sum_{k \in \mathbb{N}} \frac{e^{-zm}}{m^2} (\varepsilon m)^{2k+1}$$

which is divergent for all  $\varepsilon > 0$  since  $\varepsilon \geq 1/m$  for  $m$  large enough.

We see that divergence comes from coefficients arising with successive differentiations.

On the other hand, if one considers the transformation generated by the truncated Hamiltonian  $\sum_{n \geq 1} \varepsilon^n X_n(\theta_1, \theta_2)$  then the transformed Hamiltonian becomes

$$\begin{aligned}\mathcal{H}(I_1, I_2; \theta_1, \theta_2) &= I_1 - \varepsilon I_2 \\ &\quad + \varepsilon^{N+1} \partial_{\theta_2} X_N(\theta_1, \theta_2),\end{aligned}$$

where  $\partial_{\theta_2} X_N(\theta_1, \theta_2) = \cos(\theta_1 - N\frac{\pi}{2}) f^{(N)}(\theta_2)$ .

Especially, with  $g(z) = 1/\exp(z) - 1$ , we have:

$$\begin{aligned}\partial_{\theta_2} X_{2N}(\theta_1, \theta_2) \\ = -\cos(\theta_1) \operatorname{Re}\left(g^{(2N-1)}(-\alpha + i\theta_2)\right)\end{aligned}$$

for  $N > 0$ ,

$$\partial_{\theta_2} X_{2N+1}(\theta_1, \theta_2) = \sin(\theta_1) \operatorname{Im}\left(g^{(2N)}(-\alpha + i\theta_2)\right)$$

for  $N > 0$ ,

Now, around the real axis, the main term in the asymptotic expansion of  $g^{(n)}(z)$  as  $n$  goes to

infinity is given by the derivative of the polar term  $1/z$  in the Laurent expansion of  $g$  at 0.

Indeed, we consider the function  $h(z) = g(z) - 1/z$  which is real analytic and admits an analyticity width of size  $2\pi$  around the real axis. Hence, Cauchy estimates ensure that for  $n$  large enough and  $z$  in the strip of width  $p$  around the real axis, the derivative  $h^{(n)}(z)$  becomes negligible with respect to the derivative of  $1/z$ .

Consequently,  $g^{(n)}(z)$  is *equivalent* to  $(-1)^n n! / z^{n+1}$  as  $n$  goes to infinity for  $z$  in the strip of width  $p$  around the real axis.

Hence, the remainder  $\partial_{\theta_2} X_N(\theta_1, \theta_2)$  has a size of order  $(N-1)! / \alpha^N$  for  $N$  large. This latter estimate *cannot be improved*, since  $\partial_{\theta_2} X_{2N}(\pi, 0)$  and  $\partial_{\theta_2} X_{2N+1}(\pi/2, \alpha \tan(\pi/4N + 2))$  admit the same size of order  $(N-1)! / \alpha^N$  for  $N$  large.

Now, we can make a “summation at the smallest term” as in section “The Case of a Single Frequency System” to obtain an optimal normalization with an *exponentially small* remainder.

## Bibliography

### Primary Literature

- Arnold VI (1964) Instability of dynamical systems with several degrees of freedom. *Sov Math Dokl* 5:581–585
- Arnold VI, Kozlov VV, Neishtadt AI (2006) Mathematical aspects of classical and celestial mechanics. In: *Encyclopaedia of mathematical sciences* 3. Dynamical systems 3, 3rd edn. Springer, New York
- Bambusi D (1999a) Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations. *Math Z* 230(2):345–387
- Bambusi D (1999b) On long time stability in Hamiltonian perturbations of non-resonant linear PDEs. *Nonlinearity* 12(4):823–850
- Bambusi D, Giorgilli A (1993) Exponential stability of states close to resonance in infinite-dimensional Hamiltonian systems. *J Stat Phys* 71(3–4):569–606
- Bambusi D, Grébert B (2006) Birkhoff normal form for partial differential equations with tame modulus. *Duke Math J* 135(3):507–567
- Bambusi D, Nekhoroshev NN (2002) Long time stability in perturbations of completely resonant PDE's. *Acta Appl Math* 70(1–3):1–22
- Benettin G (2005) Physical applications of Nekhoroshev theorem and exponential estimates. In: Giorgilli A (ed) *Cetraro (2000) Hamiltonian dynamics, theory and applications*. Springer, New York
- Benettin G, Fasso F (1996) Fast rotations of the rigid body: a study by Hamiltonian perturbation theory, I. *Nonlinearity* 9(1):137–186



- Benettin G, Gallavotti G (1986) Stability of motions near resonances in quasi-integrable Hamiltonian systems. *J Stat Phys* 44(3–4):293–338
- Benettin G, Giorgilli A (1994) On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. *J Stat Phys* 74(5–6):1117–1143
- Benettin G, Sempio P (1994) Adiabatic invariants and trapping of a point charge in a strong non-uniform magnetic field. *Nonlinearity* 7(1):281–303
- Benettin G, Galgani L, Giorgilli A (1985) A proof of Nekhoroshev's theorem for the stability times in nearly-integrable Hamiltonian systems. *Celest Mech* 37:1–25
- Benettin G, Fröhlich J, Giorgilli A (1988) A Nekhoroshev-type theorem for Hamiltonian systems with infinitely many degrees of freedom. *Commun Math Phys* 119(1):95–108
- Benettin G, Fassò F, Guzzo M (1998) Nekhoroshev stability of L4 and L5 in the spatial restricted three body problem. *Regul Chaotic Dyn* 3(3):56–72
- Blaom AD (2001) A geometric setting for Hamiltonian perturbation theory. *Mem Am Math Soc* 727(xviii):112
- Bogolyubov NN, Mitropol'skij YA (1958) Asymptotic methods in the theory of nonlinear oscillations, 2nd edn. Nauka, Moscow. Engl. Transl. (1961) Gordon and Breach, New York
- Bourgain J (2004) Remarks on stability and diffusion in high-dimensional Hamiltonian systems and partial differential equations. *Ergod Theory Dyn Syst* 24(5):1331–1357
- Carati A, Galgani L, Giorgilli A, Ponno A (2002) The Fermi–Pasta–Ulam problem. *Nuovo Cimento B* 117:1017–1026
- Carati A, Galgani L, Giorgilli A (2006) Dynamical systems and thermodynamics. In: Françoise JP, Naber GL, Tsun TS (eds) *Encyclopedia of mathematical physics*. Elsevier, Amsterdam
- Celletti A, Giorgilli A (1991) On the stability of the Lagrangian points in the spatial restricted problem of three bodies. *Celest Mech Dyn Astron* 50(1):31–58
- Chirikov BV (1979) A universal instability in many dimensional oscillator systems. *Phys Rep* 52:263–279
- Delshams A, Gutierrez P (1996) Effective stability and KAM theory. *J Differ Equ* 128(2):415–490
- Dullin H, Fassò F (2004) An algorithm for detecting directional quasi-convexity. *BIT* 44(3):571–584
- Dumas HS (1993) A Nekhoroshev-like theory of classical particle channeling in perfect crystals. In: Jones CKRT et al (eds) *Dynamics reported. Expositions in dynamical systems, new series, vol 2*. Springer, Berlin
- Dumas HS (2005) Mathematical theories of classical particle channeling in perfect crystals. *Nucl Inst Meth Phys Res Sect B (Beam Interactions with Materials and Atoms)* 234(1–2):3–13
- Efthymiopoulos C, Giorgilli A, Contopoulos G (2004) Nonconvergence of formal integrals. II: Improved estimates for the optimal order of truncation. *J Phys A Math Gen* 37(45):10831–10858
- Fassò F, Lewis D (2001) Stability properties of the Riemann ellipsoids. *Arch Ration Mech Anal* 158(4):259–292
- Fassò F, Guzzo M, Benettin G (1998) Nekhoroshev-stability of elliptic equilibria of Hamiltonian systems. *Commun Math Phys* 197(2):347–360
- Giorgilli A (1998) On the problem of stability for near to integrable Hamiltonian systems. In: *Proceedings of the international congress of Mathematicians, Berlin 1998*. Documenta Mathematica III, pp 143–152
- Giorgilli A, Morbidelli A (1997) Invariant KAM tori and global stability for Hamiltonian systems. *Z Angew Math Phys* 48(1):102–134
- Giorgilli A, Skokos C (1997) On the stability of the Trojan asteroids. *Astron Astroph* 317:254–261
- Giorgilli A, Zehnder E (1992) Exponential stability for time dependent potential. *ZAMP* 43:827–855
- Giorgilli A, Delshams A, Fontich E, Galgani L, Simó C (1989) Effective stability for a Hamiltonian system near an elliptic equilibrium point, with an application to the restricted three bodies problem. *J Differ Equ* 77:167–198
- Gramchev T, Popov G (1995) Nekhoroshev type estimates for billiard ball maps. *Ann Inst Fourier* 45(3):859–895
- Guzzo M (2003) Nekhoroshev stability of asteroids. Triennial report 2000–2003 of Commission 7-Celestial Mechanics and dynamical Astronomy of the IAU; Reports on Astronomy, 1999–2002. Transactions of the International Astronomical Union, vol XXVA, Astronomical Society of the Pacific
- Guzzo M (2004) A direct proof of the Nekhoroshev theorem for nearly integrable symplectic maps. *Ann Henri Poincaré* 5(6):1013–1039
- Guzzo M, Benettin G (2001) A spectral formulation of the Nekhoroshev theorem and its relevance for numerical and experimental data analysis. *Discret Contin Dyn Syst Ser B* 1(1):1–28
- Guzzo M, Morbidelli A (1997) Construction of a Nekhoroshev-like result for the asteroid belt dynamical system *Celest. Mech Dyn Astron* 66:255–292
- Guzzo M, Lega E, Froeschlé C (2006) Diffusion and stability in perturbed non-convex integrable systems. *Nonlinearity* 19(5):1049–1067
- Hairer E, Lubich C, Wanner G (2006) Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations. Springer Series in Computational Mathematics, 2nd edn, vol 31. Springer, New York
- Ilyashenko IS (1986) A steepness test for analytic functions. *Russ Math Surv* 41:229–230
- Kuksin SB (2006) Hamiltonian PDEs (with an appendix by Dario Bambusi). In: Hasselblatt B (ed) *Handbook of dynamical systems, vol 1B*. Elsevier, Amsterdam
- Kuksin S, Pöschel J (1994) On the inclusion of analytic symplectic maps in analytic Hamiltonian flows and its applications. In: Kuksin S, Lazutkin VF, Pöschel J (eds) *Proceedings of the 1991 Euler Institute conference on dynamical systems. Prog nonlinear differential equation with application (12)*. Birkhäuser, Basel
- Littlewood JE (1959) On the equilateral configuration in the restricted problem of the three bodies. *Proc Lond Math Soc* 9(3):343–372



- Lochak P (1992) Canonical perturbation theory via simultaneous approximation. *Russ Math Surv* 47:57–133
- Lochak P (1993) Hamiltonian perturbation theory: periodic orbits, resonances and intermittency. *Nonlinearity* 6(6): 885–904
- Lochak P (1995) Stability of Hamiltonian systems over exponentially long times: the near-linear case. In: Dumas HS, Meyer K, Schmidt D (eds) *Hamiltonian dynamical systems – history, theory and applications*. IMA conference proceedings series, vol 63. Springer, New York, pp 221–229
- Lochak P, Meunier C (1988) *Multiphase averaging methods for Hamiltonian systems*. Applied Mathematical Science series, vol 72. Springer, New York
- Lochak P, Neishtadt AI (1992) Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian. *Chaos* 2(4):495–499
- Lochak P, Neishtadt AI, Niederman L (1994) Stability of nearly integrable convex Hamiltonian systems over exponentially long times. In: Kuksin S, Lazutkin VF, Pöschel J (eds) *Proceedings of the 1991 Euler Institute Conference on Dynamical Systems*. Prog nonlinear differential equation with application (12). Birkhäuser, Basel
- Marco JP, Lochak P (2005) Diffusion times and stability exponents for nearly integrable analytic systems. *Central Eur J Math* 3(3):342–397
- Marco JP, Sauzin D (2003) Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems. *Publ Math Inst Hautes Etudes Sci* 96: 199–275
- Meyer KR, Hall GR (1992) *Introduction to Hamiltonian dynamical systems and the N-Body problem*. Applied mathematical sciences, vol 90. Springer, New York
- Moan PC (2004) On the KAM and Nekhoroshev theorems for symplectic integrators and implications for error growth. *Nonlinearity* 17(1):67–83
- Morbiddelli A, Giorgilli A (1995) Superexponential stability of KAM tori. *J Stat Phys* 78:1607–1617
- Morbiddelli A, Guzzo M (1997) The Nekhoroshev theorem and the asteroid belt dynamical system. *Celest Mech Dyn Astron* 65(1–2):107–136
- Moser J (1955) Stabilitätsverhalten Kanonischer Differentialgleichungssysteme. *Nachr Akad Wiss Göttingen, Math Phys Kl IIa* 6:87–120
- Neishtadt AI (1981) On the accuracy of conservation of the adiabatic invariant. *Prikl Mat Mekh* 45:80–87. Translated in *J Appl Math Mech* 45:58–63. Neishtadt AI (1984) The separation of motions in systems with rapidly rotating phase. *J Appl Math Mech* 48:133–139
- Nekhoroshev NN (1973) Stable lower estimates for smooth mappings and for gradients of smooth functions. *Math USSR Sb* 19(3):425–467
- Nekhoroshev NN (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russ Math Surv* 32:1–65
- Nekhoroshev NN (1979) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems 2. *Trudy Sem Petrovs* 5:5–50. Translated In: Oleinik OA (ed) *Topics in Modern Mathematics*. Petrovskii Semin, vol 5. Consultant Bureau, New York
- Niederman L (1996) Stability over exponentially long times in the planetary problem. *Nonlinearity* 9(6): 1703–1751
- Niederman L (1998) Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system. *Nonlinearity* 11:1465–1479
- Niederman L (2004) Exponential stability for small perturbations of steep integrable Hamiltonian systems. *Erg Theor Dyn Syst* 24(2):593–608
- Niederman L (2006) Hamiltonian stability and subanalytic geometry. *Ann Inst Fourier* 56(3):795–813
- Niederman L (2007) Prevalence of exponential stability among nearly integrable Hamiltonian systems. *Erg Theor Dyn Syst* 27(3):905–928. Northrop TG (1963) *The adiabatic motion of charged particles*. Interscience Publishers, New York
- Poincaré H (1892) *Méthodes Nouvelles de la Mécanique Céleste*, vol 4. Blanchard, Paris
- Pöschel J (1993) Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math Z* 213:187–217
- Pöschel J (1999) On Nekhoroshev's estimate at an elliptic equilibrium. *Int Math Res Not* 1999(4):203–215
- Ramis JP, Schäfke R (1996) Gevrey separation of fast and slow variables. *Nonlinearity* 9(2):353–384
- Steichen D, Giorgilli A (1998) Long time stability for the main problem of artificial satellites. *Cel Mech* 69: 317–330
- Treschev DV (1994) Continuous averaging in Hamiltonian systems. In: Kuksin S, Lazutkin VF, Pöschel J (eds) *Proceedings of the 1991 Euler Institute conference on dynamical systems*. Prog nonlinear differential equation with application (12). Birkhäuser, Basel
- Valdinoci E (2000) Estimates for non-resonant normal forms in Hamiltonian perturbation theory. *J Stat Phys* 101(3–4):905–919

## Books and Reviews

- Arnold VI (1983) *Geometrical methods in the theory of ordinary differential equations*. Transl. from the Russian by Joseph Szuecs, Mark Levi (ed) *Grundlehren der Mathematischen Wissenschaften* 250. Springer, New York
- Arnold VI, Kozlov VV, Neishtadt AI (2006) *Mathematical aspects of classical and celestial mechanics*. *Encyclopaedia of Mathematical Sciences* 3. Dynamical systems 3, 3rd revised edn. Springer, New York
- Benettin G (2005) Physical applications of Nekhoroshev theorem and exponential estimates. In: Giorgilli A (ed) *Cetraro (2000) Hamiltonian dynamics, theory and applications*. Springer, New York
- Giorgilli A (2003) Exponential stability of Hamiltonian systems. *Dynamical systems, Part I. Hamiltonian systems and celestial mechanics*. Selected papers from the Research Trimester held in Pisa, Italy, February 4–April 26, 2002. Pisa: Scuola Normale Superiore. Pubblicazioni del Centro di Ricerca Matematica Ennio de Giorgi. Proceedings, 87–198
- Sanders JA, Verhulst F, Murdock J (2007) *Averaging methods in nonlinear dynamical systems*, 2nd edn. Applied mathematical sciences 59. Springer, New York

# Perturbation of Superintegrable Hamiltonian Systems

Francesco Fassò  
Dipartimento di Matematica “Tullio Levi-Civita”,  
Università di Padova, Padova, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Superintegrable Hamiltonian Systems  
Examples  
Perturbations of Superintegrable Systems  
Conclusion  
Appendix: Semiglobal Normal Forms  
Bibliography

## Glossary

**Action-Angle Coordinates** In a  $2d$ -dimensional symplectic manifold  $(M, \Omega)$ , a diffeomorphism  $A = (a, \alpha) : U \rightarrow A \times \mathbb{T}^d$  from an open subset  $U$  of  $M$  into the product of an open set  $A \subseteq \mathbb{R}^d$  and the  $d$ -dimensional torus, which is such that  $\Omega|_U = \sum_{i=1}^d da_i \wedge d\alpha_i$ .

**Completely Integrable System** A Hamiltonian system on a  $2d$ -dimensional symplectic manifold which possesses  $d$  independent first integrals in involution. The *Liouville-Arnold* theorem states that if the level sets of these integrals are compact and connected, then (1) they are diffeomorphic to  $\mathbb{T}^d$ , (2) in a neighborhood of each of them, there exists a system of action-angle coordinates  $(a, \alpha) \in \mathbb{R}^d \times \mathbb{T}^d$  in which the level sets of the first integrals are the sets  $a = \text{const}$ , and (3) the Hamiltonian is independent of

the angles  $\alpha$ . Written in action-angle coordinates, a completely integrable system has Hamiltonian  $h = h(a)$ . Hamilton equations give  $\dot{\alpha} = \frac{\partial h}{\partial a}(a)$ ,  $\dot{a} = -\frac{\partial h}{\partial \alpha} = 0$ , and the flow is linear on each torus  $a = \text{const}$  with frequency vector  $\omega(a)$  given by the *frequency map*  $\omega : \mathbb{R}^d \rightarrow \mathbb{R}^d, \alpha \mapsto \omega(a) := \frac{\partial h}{\partial a}(a)$ .

**Convexity and Quasi-Convexity** A completely integrable Hamiltonian with  $d$  degrees of freedom written in action-angle coordinates,  $h = h(a)$ , with frequency map  $\omega = \frac{\partial h}{\partial a}$ , is said to be *convex* if, at each point  $a$ ,  $v \cdot \frac{\partial \omega}{\partial a}(a)v = 0$  for  $v \in \mathbb{R}^d$  implies  $v = 0$  and *quasi-convex* if, at each point  $a$ ,  $v \cdot \frac{\partial \omega}{\partial a}(a)v = 0$  and  $v \cdot \omega(a) = 0$  with  $v \in \mathbb{R}^d$  imply  $v = 0$ . Convexity implies quasi-convexity which implies Kolmogorov's nondegeneracy.

**Diophantine Frequencies** A vector  $\omega \in \mathbb{R}^n$ ,  $n \geq 2$ , is called *Diophantine* if there exists  $\tau, c > 0$  such that  $|\omega \cdot v| > \frac{c}{|v|^\tau}$  for all  $v \in \mathbb{Z}^n \setminus \{0\}$ .

**Ehresmann fibration theorem** It states that every submersion with compact and connected fibers is a fibration.

**Fibration** Informally: a decomposition of a manifold into (embedded) submanifolds all diffeomorphic to each other and well arranged, given by the level sets of a map with good properties. Formally: a *(locally trivial) fibration* with total space a manifold  $M$  and base a manifold  $B$  is a surjective submersion  $y : M \rightarrow B$  with the following property: For each point  $b \in B$ , there are a neighborhood  $U \subseteq B$  and a diffeomorphism  $\mathcal{T} : y^{-1}(U) \rightarrow U \times y^{-1}(b)$ , called a *local trivialization*, such that  $y|_{y^{-1}(U)} = \pi_1 \circ \mathcal{T}$  with  $\pi_1 : U \times y^{-1}(b) \rightarrow U$  the projection onto the first factor. If  $B$  is connected, then all the fibers (=level sets) of  $y$  are diffeomorphic to each other and it is possible and common to regard the local trivializations as maps  $U \rightarrow U \times F$ , where  $F$ , called the *typical fiber*, is a manifold diffeomorphic to the fibers of  $\mathcal{T}$ . It is also possible to regard each local trivialization as a map  $U \rightarrow \psi(U) \times F$ , where  $\psi$  is a diffeomorphism, e.g., a coordinate

system for  $U$ . (This is in fact what we will do with the “generalized action-angle coordinates.”)

**Foliation** Informally: a decomposition of a manifold into (immersed) submanifolds which have the same dimension and are only locally well arranged. Formally: a (regular) foliation of rank  $r$  of a manifold  $M$  ( $1 \leq r \leq \dim M$ ) is given by a family of immersed, connected, and mutually disjoint  $r$ -dimensional submanifolds  $F_\lambda$  of  $M$ , called the leaves of the foliation, with  $\lambda$  in some index set, and by an atlas of  $M$  with charts  $\varphi : U \rightarrow \mathbb{R}^r \times \mathbb{R}^{\dim M - r} \ni (x, y)$  such that, for each  $\lambda$ , if  $U \cap F_\lambda \neq \emptyset$  then  $\varphi(U \cap F_\lambda)$  is given by  $y = \text{const}$ . The leaves of a foliation need not be level sets of submersions.

**Invariant torus** of a vector field  $X$  on a manifold  $M$  An invariant submanifold  $N \subset M$  which is diffeomorphic to  $\mathbb{T}^n$  for some  $n \geq 1$  and is such that the restriction to  $N$  of the flow of  $X$  is conjugate to a linear flow on  $\mathbb{T}^n$ .

**Kolmogorov nondegeneracy condition** Kolmogorov nondegeneracy condition for a completely integrable Hamiltonian written in action-angle coordinates,  $h = h(a)$ , is that the frequency map  $a \mapsto \omega(a) = \frac{\partial h}{\partial a}(a)$  is a local diffeomorphism, or equivalently  $\det\left(\frac{\partial^2 h}{\partial a \partial a}(a)\right) = \det\left(\frac{\partial \omega(a)}{\partial a}(a)\right) \neq 0$  for all  $a$ .

**Linear Flow on the Torus** The flow of the constant vector field  $\sum_{j=1}^n \omega_j \partial \alpha_j$  on the  $n$ -dimensional torus  $\mathbb{T}^n = (\mathbb{R}/2\pi\mathbb{Z})^n$ , namely  $(t, \alpha) \mapsto \alpha + \omega t \pmod{2\pi}$ . The vector  $\omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n$  is called the frequency vector.

**Poisson Manifold** A manifold  $P$  equipped with a Poisson bracket, namely, a bilinear, antisymmetric map  $\{, \}_P : C^\infty(P) \times C^\infty(P) \rightarrow C^\infty(P)$  which satisfies the Leibniz rule and the Jacobi identity ( $\{fg, h\}_P = f\{g, h\}_P + g\{f, h\}_P$  and  $\{f, \{g, h\}_P\}_P + \text{cyclic permutations} = 0 \forall f, g, h \in C^\infty(P)$ ). Every symplectic manifold is a Poisson manifold with its own Poisson bracket, but the bracket of a Poisson manifold may be degenerate. The *Hamiltonian vector field*  $X_f^P$  of a function  $f \in C^\infty(P)$  is the unique vector field on  $P$  such that  $L_{X_f^P} g = -\{f, g\}_P \forall g \in C^\infty(P)$ . Functions defined in (open subsets of)  $P$  which Poisson-commute

with all other functions are called (local) *Casimirs*;  $f$  is a Casimir if and only if  $X_f^P = 0$ . The subspaces  $\mathcal{H}_p := \{X_f^P(p) : f \in C^\infty(P)\}$ ,  $p \in P$ , form the so-called characteristic distribution of the Poisson manifold, which is Frobenius integrable and whose integral manifolds are symplectic manifolds. We consider only the case in which the characteristic distribution has constant rank, namely  $\dim \mathcal{H}_p = 2r \forall p \in P$  with some  $1 \leq r \leq \frac{1}{2} \dim P$ . Then, its  $2r$ -dimensional integral manifolds are the leaves of a (regular) foliation  $\mathcal{S}$  of  $P$ , called the *symplectic foliation* of  $P$ . Each leaf  $S$  of  $\mathcal{S}$  carries a symplectic form  $\Omega_S$  which is “compatible” with the Poisson structure of  $P$ , in the sense that the inclusion  $\iota : S \rightarrow P$  is a Poisson morphism between  $S$  equipped with the Poisson structure  $\{, \}_S$  induced by  $\Omega_S$  and  $(P, \{, \}_P)$ , namely  $\{f \circ \iota, g \circ \iota\}_S(s) = \{f, g\}_P(\iota(s))$  for all  $s \in S$  and  $f, g \in C^\infty(P)$ . The *rank* of  $(P, \{, \}_P)$  is the dimension of the leaves of its symplectic foliation. The number of functionally independent (local) Casimirs equals  $\dim P - 2r$ .

**Resonances and Resonant Lattice** A linear flow on  $\mathbb{T}^n$  and its frequency vector  $\omega$  are said to be *resonant* (resp. *nonresonant*) if there exists a (resp. there exists no) nonzero integer vector  $v \in \mathbb{Z}^n$ , also called a *resonance* of  $\omega$ , such that  $\omega \cdot v = 0$ . The  $L^1$ -norm  $|v|$  of the resonance  $v$  is called its *order*. The set of resonances of a given vector  $\omega \in \mathbb{R}^n$  forms a lattice  $L \subseteq \mathbb{Z}^n$  of some rank  $r$ ,  $1 \leq r \leq n$  (namely, the set of all the linear combinations with integer coefficients of  $r$  linearly independent vectors  $v_1, \dots, v_r \in \mathbb{Z}^n$ , called a basis of  $L$ ). If  $\omega$  is nonresonant, then each orbit  $(t, \alpha) \mapsto \alpha + \omega t \pmod{2\pi}$  is dense in  $\mathbb{T}^n$ ; if the resonant lattice of  $\omega$  has rank  $r$ , then the closure of each such orbit is a subtorus of dimension  $n - r$  (Kronecker theorem).

**Steepness** Steepness of a completely integrable Hamiltonian  $h = h(a)$  written in action-angle coordinates is a condition introduced by Nekhoroshev in Nekhoroshev (1977) which generalizes convexity and quasi-convexity.

**Superintegrable Systems** Informally, a Hamiltonian system with  $d$  degrees of freedom which, due to the existence of more than

$d$  integrals of motion with suitable properties, has the flow linear on tori of dimension  $n < d$ . A precise description of these systems, and the exact relationship with complete integrability, is described in this item.

**Symplectic manifold** An even-dimensional manifold  $M$  equipped with a symplectic form, namely, a closed, nondegenerate differential 2-form  $\Omega$ . By Darboux theorem, in a neighborhood  $U$  of every point of  $M$  there exist “Darboux” coordinates  $\mathcal{D} = (q_1, \dots, q_d, p_1, \dots, p_d)$ ,  $2d = \dim M$ , such that  $\mathcal{D}^*\Omega = \sum_{i=1}^d dp_i \wedge dq_i$ . The *Hamiltonian vector field*  $X_f$  of a function  $f \in C^\infty(M)$  is the unique vector field on  $M$  such that  $i_{X_f}\Omega = -df$ , or  $\Omega(X_f, Y) = L_Y f$  for all vector fields  $Y$  on  $M$ . ( $C^\infty(M) := C^\infty(M, \mathbb{R})$  and  $L_Y f = i_Y df$  is the Lie derivative of a function  $f$  along a vector field  $Y$ .) In Darboux coordinates,  $X_h = \sum_{i=1}^d \left( \frac{\partial h}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial h}{\partial q_i} \frac{\partial}{\partial p_i} \right)$ . *Hamiltonian system*  $(M, \Omega, f)$  is a synonym for  $X_f$ , and  $d = \frac{1}{2} \dim M$  is its *number of degrees of freedom*. The *Poisson bracket*  $\{, \}_\Omega$  of  $(M, \Omega)$  associates to each pair of functions  $f, g$  the function  $\{f, g\}_\Omega$  defined by  $\{f, g\}_\Omega := -\Omega(X_f, X_g) = -L_{X_f} g$  and satisfies  $[X_f, X_g] = -X_{\{f, g\}_\Omega}$  for all functions  $f, g$ . ( $[X, Y]$  is the Lie bracket of two vector fields  $X, Y$ .) In Darboux coordinates,  $\{f, g\} = \sum_{i=1}^d \left( \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right)$ . The nondegeneracy of  $\Omega$  implies that, in a symplectic manifold, the pairing  $df \rightarrow X_f$  and the Poisson bracket are nondegenerate: If, at a point  $m \in M$ ,  $df(m) \neq 0$ , then  $X_f(m) \neq 0$  and there exists a function  $g$  such that  $\{f, g\}_\Omega(m) \neq 0$ . Two functions  $f$  and  $g$  on  $M$  are said to be *in involution* if  $\{f, g\}_\Omega = 0$ .

## Definition of the Subject

For a Hamiltonian system with  $d$  degrees of freedom, complete integrability (namely, the existence of  $d$  independent first integrals in involution) implies that (a part of interest of) the phase space is foliated by  $d$ -dimensional invariant tori on which the dynamics is linear. Complete

integrability is often regarded as the ultimate integrability notion for Hamiltonian systems. However, several of the most important integrable systems of Classical Mechanics have more first integrals than the number  $d$  of degrees of freedom, and their phase space is foliated by invariant tori of dimension  $n < d$ . For instance:

- For a point in a central force field and for the Euler top,  $d = 3$ , but there are four integrals of motion (the energy and the three components of the angular momentum vector), and motions are linear on tori of dimension  $n = 2$ .
- In the particular case of the Newtonian central force field, also known as Kepler system, there is an additional first integral which comes from the Laplace-Runge-Lenz vector and the dynamics is periodic, thus linear on tori of dimension  $n = 1$ .
- All cases with  $1 \leq n < d$  are met with resonant oscillators.

Systems of this type are called *superintegrable*. It is often possible to describe, *locally*, a superintegrable system with  $d$  degrees of freedom as a completely integrable one whose Hamiltonian depends on only  $n < d$  of the actions (and hence is “degenerate”), but this description fails identifying some essential aspects of the superintegrable system.

The dynamics of small perturbations of superintegrable systems is of great importance in Classical Mechanics. In particular, the study of perturbations of a specific superintegrable system – the planetary system – has been one of the motivation for the development of Hamiltonian perturbation theory and, in particular, of the two main branches of today’s theory, namely, KAM and Nekhoroshev theories. A formulation of KAM theorem that applies to perturbations of superintegrable systems of the type met in the planetary system was given by Arnold in the very early stages of the development of KAM theory (Arnold 1963). It rests on the possibility of using part of the perturbation to build, at least in a subset of interest of the phase space, an “intermediate” completely integrable system to which KAM theorem can then be applied, a technique

known as “removal of the degeneracy.” And Nekhoroshev’s “main theorem” in his celebrated 1977 article (Nekhoroshev 1977) was formulated so as to be applicable to some superintegrable cases, including planetary systems. However, in general, the applicability of Nekhoroshev theorem to superintegrable systems faces a difficulty that comes from the fact that these systems typically do not possess global sets of coordinates of action-angle type, and the formulation of Nekhoroshev theorem requires certain “semiglobal” considerations that rest on the symplectic geometry of the superintegrable systems.

In short, in the phase space of a superintegrable system there are *two* naturally defined invariant foliations. One is the foliation by the invariant  $n$ -dimensional tori. The leaves of the other foliation have dimension  $2d - n > d$  and are unions of tori carrying motions with the same frequencies. The normal forms that enter the proof of Nekhoroshev theorem need to be constructed “semiglobally,” in neighborhoods of the leaves of the latter foliation, which may have (and often have) a nontrivial geometry and cannot be covered with global coordinates. The very possibility of building these “semiglobal” normal forms requires some knowledge of the symplectic geometry of the bifoliated structure of the phase space of the superintegrable systems.

## Introduction

**History** Superintegrable systems have attracted attention throughout the entire development of classical mechanics. The work of Newton on Kepler system, that of Euler on the rigid body, and Bertrand theorem on central force fields, for instance, all pertain to superintegrability. A good source for what was classically known is (Born 1960). A more recent milestone was the discovery by Fock of the  $SO(4)$ -symmetry of the hydrogen atom in quantum mechanics, later extended to the Kepler system in classical mechanics. The field has undergone a great development in the last fifty years, along two somewhat different directions.

On the one hand, starting with the work of Frisch, Smorodinskii, Winternitz, and others in the 1960s, an imponent amount of work has been dedicated to the case of *maximally superintegrable* systems, which have  $2d - 1$  first integrals, with  $d$  the number of degrees of freedom, and periodic flow. This part of the literature has focused on problems such as the construction of potentials producing maximally superintegrable systems, the connection with the separability of the Hamilton-Jacobi equation and with the spectrum of the Schrödinger operator, etc. We refer, e.g., to (Evans 1990; Kibler and Winternitz 1990; Miller et al. 2013; Tempesta et al. 2004) where descriptions of these researches and references to the extremely vast literature can be found.

On the other hand, a comprehensive study of superintegrable systems, with motions on tori of *any* dimension  $n$  between 1 and  $d - 1$ , began with the work of Nekhoroshev in 1972 (Nekhoroshev 1972), which was later extended by Mischenko and Fomenko’s “noncommutative integrability” (Mischenko and Fomenko 1978). These studies clarified the properties of the  $2d - n$  first integrals that lead to superintegrability and the geometric structure of the fibration by the invariant tori. In particular, the global geometry of this fibration in the superintegrable case was studied by Nekhoroshev (Nekhoroshev 1972) and Dazord and Delzant (Dazord and Delzant 1987) (the study of the completely integrable case is due to Duistermaat (Duistermaat 1980)). The characterization in terms of bifibration and dual pairs is present there and in the book (Karasev and Maslov 1993). It is this approach which is relevant for perturbation theory and will be covered in this item.

The formulation of Nekhoroshev theorem for superintegrable system is due to (Fassò 1995) and (Blaom 2001). Some applications will be described below. A review of these topics is (Fassò 2005).

**Organization of the Entry** The readers of this entry are assumed to have some knowledge of complete integrability and of KAM and Nekhoroshev theories and to be interested in the aspects of these theories related to the superintegrability of the unperturbed system. (For

general introductions to these theories, see Chap. 6 of Arnold et al. (2006), Benettin (2005); Sevryuk (2003) and the entries (Broer and Hanßmann; Chierchia and Procesi; Niederman) in this Encyclopedia). Since, as said above, this requires a comprehension of the much less known geometric structure of the fibration by the invariant tori of superintegrable systems, a first part of the present entry is devoted to it. In this first part, we describe some integrability notions for Hamiltonian systems and the structure of the superintegrable ones, and we give some examples. Subsequently, we review the application of Nekhoroshev and KAM theories to superintegrable systems.

**Conventions** All objects (manifolds, functions, vector fields, etc.) will be supposed to be smooth.  $L_X f = i_X df$  denotes the Lie derivative of a function  $f$  along a vector field  $X$ ,  $[X, Y]$  the Lie bracket of two vector fields  $X, Y$ , and  $\Phi_t^X$  the map at time  $t$  of the flow  $\Phi^X$  of a vector field  $X$ .

**Remark** Typically, the foliation by the invariant tori of an integrable system has singularities (e.g., lower dimensional tori and/or, with the terminology that we will introduce, singularities of the symplectic leaves and of the flowers, etc.) whose consideration may have relevant interest from a perturbation point of view. Nevertheless, with the exception of one example, we will restrict our attention to the regular part of the phase space, where all invariant tori have the same dimension and form a fibration.

## Superintegrable Hamiltonian Systems

### A Paradigm for Integrability

The notion of integrability appropriate to classical mechanics refers to the quasi-periodicity of the dynamics: The phase space (or a subset of interest of it, to which the consideration is restricted) is foliated by submanifolds which are diffeomorphic to the torus  $\mathbb{T}^n$  for some given  $n$ , are invariant under the dynamics, and the restriction of the flow to each of them is conjugate to a linear flow on  $\mathbb{T}^n$ . Often, this foliation is a fibration. In general, the existence of this structure is due to the combined existence of

first integrals and symmetries. For Hamiltonian systems – due to their double variational and symplectic nature – first integrals and symmetries are linked to each other and the characterization of integrability can be made in terms of first integrals alone. However, in order to better understand the superintegrable Hamiltonian case, it is useful to start from the non-Hamiltonian case, with the following basic result stated around 1995 in Bogoyavlenskij (1998); Fedorov (1999); Zung (2006):

**Proposition 1** *Assume that on a manifold  $M$  there are, for some  $1 \leq n < \dim M$ ,*

- *A submersion  $y = (y_1, \dots, y_{\dim M - n}) : M \rightarrow \mathbb{R}^{\dim M - n}$  with compact and connected fibers*
- *$n$  everywhere linearly independent and pairwise commuting vector fields  $Y_1, \dots, Y_n$  which are tangent to the fibers of  $y$ :*

$$\begin{aligned} [Y_j, Y_\ell] &= 0 \quad \text{and} \quad L_{Y_{y_r}} = 0 \\ \forall j, \ell &= 1, \dots, n, \quad r = 1, \dots, \dim M - n. \end{aligned} \quad (1)$$

*Then:*

- (i) *The fibers of  $y$  are diffeomorphic to  $\mathbb{T}^n$ .*
- (ii) *Every fiber of  $y$  has a neighborhood  $U$  endowed with a diffeomorphism*

$$\begin{aligned} \mathcal{T} &= (b, \alpha) : U \rightarrow \mathcal{B} \times \mathbb{T}^n, \\ \mathcal{B} &= b(U) \subset \mathbb{R}^{\dim M - n}, \end{aligned}$$

*such that the level sets of  $y$  coincide with the level sets of  $b$  (namely,  $y|_U = \pi_1 \circ \mathcal{T}$  with  $\pi_1$  the projection onto the first factor).*

- (iii) *Assume that a vector field  $X$  on  $M$  has the  $y_r$ 's as first integrals and the  $Y_j$ 's as dynamical symmetries, namely*

$$\begin{aligned} L_X y_r &= 0 \quad \text{and} \quad [X, Y_j] = 0 \\ \forall r &= 1, \dots, \dim M - n, \quad j = 1, \dots, n. \end{aligned}$$

*Then, the fibers of  $y$  are invariant under the flow of  $X$ , and each diffeomorphism  $\mathcal{T}$  as in ii. conjugates the flow of  $X$  in  $U$  to the linear flow*



$$\Phi_t^{\mathcal{T} \circ X}(b, \alpha) = (b, \alpha + \omega(b)t \mod 2\pi)$$

on  $\mathbb{T}^n$ , with a map  $\omega : b(U) \rightarrow \mathbb{R}^n$ .

For everyone familiar with the proof of the Liouville-Arnol'd theorem, as given, for instance, in Arnold (1989), this result should appear as essentially self-evident and its proof elementary. Indeed, the fibers of  $y$  are compact and connected  $n$ -dimensional submanifolds which carry  $n$  commuting and everywhere linear independent tangent vector fields (the  $Y_j$ 's). Therefore, each fiber  $F$  of  $y$  is diffeomorphic to  $\mathbb{T}^n$ : The diffeomorphism  $\alpha : F \rightarrow \mathbb{T}^n$  is constructed by showing that, on  $F$ , there are  $n$  linear combinations  $\partial_{\alpha_1}, \dots, \partial_{\alpha_n}$  with constant coefficients of the  $Y_j$ 's whose flows are all periodic with period  $2\pi$ ; the  $\alpha_j$ 's are the times ( $\mod 2\pi$ ) along the flows of these vector fields, measured from a certain point of  $F$ . Smoothly extending this construction to the nearby fibers gives the diffeomorphism  $\mathcal{T}$ . Note that, by the Ehresmann theorem,  $y : M \rightarrow \mathbb{R}^{\dim M - n}$  is a fibration with fiber  $\mathbb{T}^n$  and the  $\mathcal{T}$ 's form an atlas of local trivializations. Finally, the vanishing of the Lie brackets  $L_X Y_r$  implies that  $X$  is tangent to the fibers of  $y$ , and the vanishing of the  $[X, Y_j]$  implies that, on each fiber of  $y$ ,  $X$  is a linear combination with constant coefficients of the  $Y_j$ 's, and hence of the  $\partial_{\alpha_j}$ 's. Thus,  $\mathcal{T}_* X(b, \alpha) = \sum_{j=1}^n \omega_j(b) \partial_{\alpha_j}$  with certain  $\omega_j(b) \in \mathbb{R}$ .

We call *invariant tori* of  $X$  the fibers of  $y$ . Note that they are generated by the flows of the commuting dynamical symmetries  $Y_1, \dots, Y_n$  (in the sense that they are the orbits of the  $\mathbb{R}^n$ -action defined by these commuting flows).

### The Hamiltonian Case

We describe now various notions of integrability of Hamiltonian systems in the light of Proposition 1. We consider a symplectic manifold  $(M, \Omega)$  of dimension  $2d$ ,  $d \geq 1$ . Recall that the Hamiltonian vector field  $X_f$  of a function  $f : M \rightarrow \mathbb{R}$  is defined by  $i_{X_f} \Omega = -df$ ; by *Hamiltonian system*  $(M, \Omega, f)$  or simply  $(M, f)$ , we mean the vector field  $X_f$ . In the following,  $\{, \}$  denotes the Poisson bracket on  $M$  induced by  $\Omega$ , which is defined by  $\{f, g\} := -\Omega(X_g, X_f)$  and satisfies  $[X_f, X_g] = -X_{\{f, g\}}$  for all functions  $f, g$ .

1. *Liouville-Arnol'd's complete integrability.* A Hamiltonian system on  $M$  is said to be *completely integrable* if it possesses  $d$  first integrals  $y_1, \dots, y_d$  which are pairwise in involution,

$$\{y_j, y_\ell\} = 0 \quad \forall j, \ell = 1, \dots, d,$$

and such that the map  $y = (y_1, \dots, y_d)$  is a submersion with compact and connected fibers. The dynamical symmetries complementary to the first integrals  $y_j$ 's are provided by their Hamiltonian vector fields  $X_{y_1}, \dots, X_{y_d}$ . In fact, these vector fields are linearly independent (because  $dy_1, \dots, dy_{2d-n}$  are linearly independent),  $L_{X_{y_\ell}} y_j = \{y_\ell, y_j\} = 0$  and  $[X_{y_j}, X_{y_\ell}] = X_{\{y_j, y_\ell\}} = 0$  for all  $j, \ell$ . Thus, the invariant tori are generated by the Hamiltonian vector fields of the first integrals in involution.

2. *Nekhoroshev's integrable degenerate systems.* In Nekhoroshev (1972), Nekhoroshev considered the case of a Hamiltonian system which, for some  $1 \leq n \leq d$ , possesses  $2d - n$  first integrals  $y_1, \dots, y_{2d-n}$  such that the first  $n$  of them are in involution with all others,

$$\begin{aligned} \{y_j, y_r\} &= 0 \quad \forall j = 1, \dots, n, \\ \tau &= 1, \dots, 2d - n, \end{aligned} \quad (2)$$

and the map  $y = (y_1, \dots, y_d)$  is a submersion with compact and connected fibers. The dynamical symmetries complementary to the first integrals are now provided by the Hamiltonian vector fields  $X_{y_1}, \dots, X_{y_n}$  of the  $n$  commuting first integrals. The invariant tori have dimension  $n$ .

3. *Mischenko and Fomenko's noncommutative integrability.* The notion of *noncommutative integrability*, introduced by Mischenko and Fomenko in Mischenko and Fomenko (1978), can be formulated in different but essentially equivalent ways. The most common one, which appears, e.g., in Fomenko (1988), is that of a Hamiltonian system which, for some  $1 \leq n \leq d$ , possesses  $2d - n$  first integrals  $y_1, \dots, y_{2d-n}$  such that the



map  $y = (y_1, \dots, y_{2d-n}) : M \rightarrow \mathbb{R}^{2d-n}$  is a submersion with compact and connected fibers and

$$\{y_r, y_s\} = P_{rs} \circ y, \quad r, s = 1, \dots, 2d-n, \quad (3)$$

with some functions  $P_{rs} : y(M) \rightarrow \mathbb{R}$  such that

$$\text{rank}(P(b)) = 2d - 2n \quad \forall b \in y(M) \quad (4)$$

where, at each point  $b \in y(M)$ ,  $P(b)$  is the (antisymmetric) matrix with entries  $P_{rs}(b)$ . This situation was called noncommutative integrability because the first integrals  $y$  form a non-abelian Lie algebra and the emphasis in Mischenko and Fomenko (1978) was on the properties of such an algebra; moreover, in natural cases, the first integrals  $(y_1, \dots, y_{2d-n})$  emerge as (energy-)momentum map of a large, nonabelian symmetry group (we will come back on this in the *Examples* section). It is not immediately clear how this situation fits in that of Proposition 1, namely, which are (possible choices of) the dynamical symmetries  $Y_1, \dots, Y_n$ . We will give an answer to this question in Proposition 5.

Given that Mischenko and Fomenko's integrability notion is the most general of the three (if  $n = d$ , then all  $P_{rs} = 0$  because  $\text{rank}(P) = 0$ ), we adopt it here as the characterization of integrability in the Hamiltonian setting. Since we also need to consider the fibration given by the first integrals per se – independently of the Hamiltonian of the system – we give it a (provisional) name as well.

**Definition 1** Let  $(M, \Omega)$  be a symplectic manifold of dimension  $2d$ ,  $d \geq 1$ .

- D1. Let  $1 \leq n \leq d$ . An integrable  $\mathbb{T}^n$ -fibration on  $M$  is a submersion  $y : M \rightarrow \mathbb{R}^{2d-n}$  which has compact and connected fibers and satisfies (3) and (4).
- D2. An integrable Hamiltonian system  $(M, h, y)$  is formed by an integrable  $\mathbb{T}^n$ -fibration of  $M$ , for some  $1 \leq n \leq d$ , and by a function  $h : M \rightarrow \mathbb{R}$  which is constant on the fibers of

$y$ , namely such that  $y_1, \dots, y_{2d-n}$  are first integrals of  $X_h$  or

$$\{h, y_r\} = 0 \quad \forall r = 1, \dots, 2d-n.$$

The fibers of  $y$  are called the invariant tori of the system.

- D3. An integrable Hamiltonian system  $(M, h, y)$  is called Liouville integrable or completely integrable if  $n = d$  and noncommutatively integrable or superintegrable if  $n < d$ .

**Remark** Hamiltonian systems with dynamics linear on invariant tori of dimension  $> n$  are possible only in very special situations (in particular, the symplectic form must be nonexact) (Fassò 1998; Fassò and Giacobbe 2002; Herman 1989; Parasyuk 1993). There is a version of KAM theorem that applies to them (Parasyuk 1984; Sevryuk 2003). We do not consider these systems here.

### The Symplectic Structure of Noncommutatively Integrable Systems

Due to their symplectic nature, integrable Hamiltonian systems have more structure than the generic case of Proposition 1. This structure, described by the Liouville-Arnol'd theorem in the completely integrable case, and by its superintegrable generalizations by Nekhoroshev and Mischenko and Fomenko, can be expressed through the existence of particular symplectic trivializations of the fibration  $y$  known as (generalized) action-angle coordinates.

**Proposition 2** (Mischenko and Fomenko 1978; Nekhoroshev 1972) Assume that  $y : M \rightarrow \mathbb{R}^{2d-n}$  is an integrable  $\mathbb{T}^n$ -fibration. Then, there exists an atlas of local trivializations  $\mathcal{T} = (b, \alpha) : U \rightarrow \mathcal{B} \times \mathbb{T}^n$  as in item ii. of Proposition 1 which are symplectic; more precisely, writing  $b = (a_1, \dots, a_n, p_1, \dots, p_{d-n}, q_1, \dots, q_{d-n})$ ,

$$\Omega|_U = \sum_{i=1}^n da_i \wedge d\alpha_i + \sum_{u=1}^{d-n} dp_u \wedge dq_u. \quad (5)$$

In each of these local trivializations, the coordinates  $a$  and the angular coordinates  $\alpha$  are called,

respectively, (local) *actions* and (local) *angles*. No unique name is given to the  $p, q$ 's, but the set of coordinates  $(a, \alpha, p, q)$  is called either *partial* or *generalized* (local) *action-angle coordinates*. A single, global system of generalized action-angle coordinates that covers the entire manifold  $M$  might not (and often does not) exist, but an atlas for  $M$  consisting of local sets of these coordinates can always be built. Such an atlas induces an atlas for the base  $B = \gamma(M) \subseteq \mathbb{R}^{2d-n}$  of the fibration, with local coordinates  $(\tilde{a}, \tilde{p}, \tilde{q})$  which are such that  $(a, p, q) = (\tilde{a}, \tilde{p}, \tilde{q}) \circ \gamma$ .

**Proposition 3** *Assume that  $(M, h, \gamma)$  is an integrable Hamiltonian system. Then, the local representative  $h^{\text{loc}}$  of  $h$  in each system of generalized action-angle coordinates  $(a, \alpha, p, q)$  is a function of the actions alone,  $h^{\text{loc}} = h^{\text{loc}}(a)$ , and Hamilton equations are*

$$\dot{a} = 0, \quad \dot{p} = 0, \quad \dot{q} = 0, \quad \dot{\alpha} = \frac{\partial h^{\text{loc}}}{\partial a}(a). \quad (6)$$

**Proof** The coordinates  $(a, p, q)$  are first integrals of the system; hence, Hamilton equations give  $\frac{\partial h^{\text{loc}}}{\partial \alpha} = \frac{\partial h^{\text{loc}}}{\partial p} = \frac{\partial h^{\text{loc}}}{\partial q} = 0$ .

Compared to the general case of Proposition 1, we encounter here a specificity of the Hamiltonian case: In each system of generalized action-angle coordinates, the frequency vector  $\omega = \frac{\partial h^{\text{loc}}}{\partial a}$  is a function only of the actions  $a$ , not of all the coordinates  $(a, p, q)$  on the base.

**Remark** The two extreme cases of Proposition 2, with  $n = d$  and  $n = 1$ , were discovered before the general case. The former is the Liouville-Arnold theorem. If  $n = 1$ , then the fibers of  $\gamma$  are one-dimensional compact manifolds – hence diffeomorphic to circles – and the flow is periodic. There is a single local action  $a$  which, if  $\frac{\partial h}{\partial a} \neq 0$ , is a locally invertible function of the Hamiltonian,  $a = a(h)$ . Hence, the period of the motions is a function of the Hamiltonian. This is *Gordon's period-energy relation*, according to which if a Hamiltonian system has periodic flow, with the periodic orbits being the fibers of a

submersion, then its period is a function of the Hamiltonian (Gordon 1969/1970).

### The Geometric Structure of Integrable Toric Fibrations

The fibration by the invariant tori of an integrable Hamiltonian system has a rich structure, which can be disclosed by investigating the transition functions between the local systems of generalized action-angle coordinates.

**Proposition 4** (Born 1960; Nekhoroshev 1972) *Let  $(a, \alpha, p, q)$  and  $(a', \alpha', p', q')$  be two sets of local generalized action-angle coordinates of an integrable  $\mathbb{T}^n$ -fibration whose domains have connected nonempty intersection. Then*

$$\begin{aligned} a' &= Z^{-T}a + z, & \alpha' &= Z\alpha + \mathcal{F}(a, p, q), \\ p' &= \mathcal{P}(a, p, q), & q' &= \mathcal{Q}(a, p, q) \end{aligned} \quad (7)$$

with some matrix  $Z \in SL_{\pm}(\mathbb{Z}, n)$ , some vector  $z \in \mathbb{R}^n$ , and some maps  $\mathcal{F}, \mathcal{P}, \mathcal{Q}$ .

**Proof** Since  $(a, p, q)$  and  $(a', p', q')$  have the same fibers (those of  $\gamma$ ), the transition functions between the two sets of coordinates have the form  $a' = a'(a, p, q)$ ,  $p' = p'(a, p, q)$ ,  $q' = q'(a, p, q)$ ,  $\alpha' = \alpha'(a, p, q, \alpha)$  and satisfy the condition of symplecticity

$$\begin{aligned} \sum_j da'_j \wedge d\alpha'_j + \sum_u dp'_u \wedge dq'_u \\ = \sum_j da_j \wedge d\alpha_j + \sum_u dp_u \wedge dq_u. \end{aligned} \quad (8)$$

The only terms coming from the l.h.s. which contain the wedge product of some  $da'_j$ 's and  $d\alpha'_j$ 's are  $\sum_{j\ell} \frac{\partial a'_j}{\partial \alpha_\ell}(a) \frac{\partial \alpha'_j}{\partial \alpha_\ell}(a, p, q, \alpha) da_h \wedge d\alpha_\ell$ . Equating this expression to  $\sum_j da_j \wedge d\alpha_j$  gives  $\frac{\partial \alpha'}{\partial \alpha}(a, p, q, \alpha) = \frac{\partial a'}{\partial a}(a)^{-T}$ . Hence,  $\frac{\partial \alpha'}{\partial \alpha}$  is independent of  $\alpha$  and there exist a matrix  $C(a)$  and a vector  $\mathcal{F}(a, p, q)$  which depend smoothly on  $(a, p, q)$  and are such that  $\alpha' = C(a)\alpha + \mathcal{F}(a, p, q) \pmod{2\pi}$ . Since  $\alpha$  and  $\alpha'$  are both angular coordinates on the tori  $f = \text{const}$ , the matrix  $C(a)$  belongs to  $SL_{\pm}(n, \mathbb{Z})$  and is constant in connected sets. Writing  $Z = C^{-T}$ , we have

$a' = Za + z(a, p, q)$  with some function  $z$ . Equating to zero all terms at the l.h.s. of (8) which contain some  $dp'_u \wedge da'_j$  or  $dq'_u \wedge da'_j$  gives  $\frac{\partial a'}{\partial p} = \frac{\partial a'}{\partial q} = 0$ . Hence  $z$  is constant.

In order to draw some consequences from this Proposition, we recall some terminology about symplectic manifolds. If  $N$  is a submanifold of a symplectic manifold  $(M, \Omega)$  and  $x \in N$ , the symplectic orthogonal  $(T_x N)^\perp$  of the tangent space  $T_x N$  is the subspace  $(T_x N)^\perp = \{u \in T_x M : \Omega(u, v) = 0 \forall v \in T_x N\}$  of  $T_x M$ . If  $N$  has dimension  $n$ , then  $(T_x N)^\perp$  has dimension  $2d - n$ . A submanifold of  $M$  is called *isotropic* (resp. *coisotropic*) if its tangent spaces are contained in (resp. contain) their own symplectic complements. Isotropic (coisotropic) submanifolds have dimension  $\leq d$  ( $\geq d$ ). *Lagrangian* submanifolds are both isotropic and coisotropic.

**Proposition 5** *Let  $y : M \rightarrow B$  be an integrable  $\mathbb{T}^n$ -fibration of a  $2d$ -dimensional symplectic manifold  $M$ . Then:*

- i. *The fibers of  $y$  are isotropic.*
- ii. *The base  $B$  has a Poisson structure of rank  $2(d - n)$ , and  $y : M \rightarrow B$  is a Poisson morphism. In the domain of any set of local coordinates  $(\tilde{a}, \tilde{p}, \tilde{q})$  on  $B$ , induced by a set of generalized action-angle coordinates  $(a, \alpha, p, q)$  on  $M$ , the symplectic leaves of  $B$  are the sets  $\tilde{a} = \text{const}$  and  $\tilde{a}_1, \dots, \tilde{a}_n$  are (local) Casimirs of  $B$ .*
- iii. *If  $(M, h, y)$  is an integrable Hamiltonian system, then the Hamiltonian vector fields of the lifts to  $M$  of the (local) Casimirs of  $B$  are dynamical symmetries of  $X_h$ .*
- iv. *There is a foliation  $C$  of  $M$  whose leaves have dimension  $2d - n$ , are union of fibers of  $y$ , are coisotropic, and are mapped by  $y$  onto the symplectic leaves of  $B$ . In each system of partial action-angle coordinates, the leaves of  $C$  are locally described by  $a = \text{const}$ .*

*Assume, moreover, that the leaves of the symplectic foliation of  $B$  are the fibers of a fibration  $\tilde{c} : B \rightarrow A$ , with  $A$  an  $n$ -dimensional manifold. Then:*

- v.  *$A$  has an integral affine structure.*

- vi. *The leaves of the foliation  $C$  are the fibers of the fibration  $c := \tilde{c} \circ y : M \rightarrow A$ .*
- vii. *If  $(M, h, y)$  is an integrable Hamiltonian system, then  $h = \hat{k} \circ c$  with a function  $\hat{k} : A \rightarrow \mathbb{R}$ .*

### Proof

- (i.) The tangent spaces to the fibers of  $y$  are spanned by the vector fields  $\partial_{x_1}, \dots, \partial_{x_n}$  and, from (5),  $\Omega(\partial_{x_j}, \partial_{x_\ell}) = 0$  for all  $j, \ell = 1, \dots, n$ .
- (ii.) For any two functions  $\tilde{f}, \tilde{g} : B \rightarrow \mathbb{R}$ , by (5),  $\{\tilde{f} \circ y, \tilde{g} \circ y\} = \sum_{rs} (\frac{\partial \tilde{f}}{\partial y_r} \circ y) (\frac{\partial \tilde{g}}{\partial y_s} \circ y) \{y_r, y_s\} = (\sum_{rs} \frac{\partial \tilde{f}}{\partial y_r} \frac{\partial \tilde{g}}{\partial y_s} P_{rs}) \circ y$ , and we may define the function  $\{f, g\}_B := \sum_{rs} \frac{\partial f}{\partial y_r} \frac{\partial g}{\partial y_s} P_{rs} : B \rightarrow \mathbb{R}$ . Thus

$$\{\tilde{f}, \tilde{g}\}_B \circ y = \{\tilde{f} \circ y, \tilde{g} \circ y\} \quad \forall \tilde{f}, \tilde{g} : B \rightarrow \mathbb{R}$$

which implies that  $\{, \}_B$  is a Poisson bracket on  $B$  and  $y$  is a Poisson morphism. In the coordinates  $\tilde{y}_r$  of  $B \subset \mathbb{R}^{2d-n}$  (which are such that  $y_r = \tilde{y}_r \circ y$ ), the matrix of the Poisson bracket  $\{, \}_B$  is  $P$  and has rank  $2(d - n)$ . (Or use the fact that, in local generalized action-angle coordinates,  $\{\tilde{p}_u, \tilde{q}_v\}_B = \{p_u, q_v\} \circ y = \delta_{uv}$ , etc.). The remaining statements are obvious, given that the dimension of the symplectic leaves equals the rank of the Poisson manifold and any maximal set of functionally independent local Casimirs has cardinality  $\dim B - \text{rank } B = n$ .

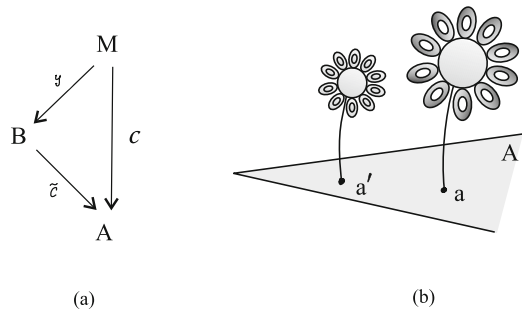
- (iii.) If  $(M, h, y)$  is an integrable Hamiltonian system, then  $h$  is constant along the fibers of  $y$  and there is a function  $\tilde{h} : B \rightarrow \mathbb{R}$  such that  $h = \tilde{h} \circ y$ . Now let  $\tilde{f} : B \rightarrow \mathbb{R}$  be a Casimir of  $B$  and  $f = \tilde{f} \circ y : M \rightarrow \mathbb{R}$  its lift. Then  $\{h, f\} = \{\tilde{h}, \tilde{f}\}_B \circ y = 0$  and therefore  $[X_f, X_h] = X_{\{h, f\}} = 0$ .
- (iv.) The first set of equations in the transition functions (8) shows that there is a foliation  $C$  of  $M$  whose leaves are locally described by  $a = \text{const}$  and hence have dimension  $2d - n$ . Since the domain of each set of partial action-angle coordinates is

saturated by the fibers of  $y$ , which are given by  $(a, p, q) = \text{const}$ , so are the leaves of  $\mathcal{C}$ . The coisotropy of the leaves of  $\mathcal{C}$  follows from the fact that, in the domain of any set of generalized action-angle coordinates, the sets  $a = \text{const}$  are symplectically orthogonal to the sets  $(a, p, q) = \text{const}$ . Finally, that  $y$  maps each leaf of  $\mathcal{C}$  onto a symplectic leaf of  $B$  follows from the fact that  $a = \tilde{a} \circ y$  and from the connectedness of the leaves of a foliation.

- (v.) The atlas of  $B$  with local coordinates  $(\tilde{a}, \tilde{p}, \tilde{q})$  induces an atlas of  $A$  with local coordinates  $\hat{a}$  such that  $\tilde{a} = \hat{a} \circ \tilde{c}$ . From (7), it follows that the transition functions between these local coordinates have the form  $\hat{a}' = Z\hat{a} + z$ , with  $Z \in SL_{\pm}(n, \mathbb{Z})$ ,  $z \in \mathbb{R}^n$ , and give  $A$  an integer affine structure.
- (vi.) By iv., the leaves of  $\mathcal{F}$  are the preimages under  $y$  of the symplectic leaves of  $B$ , which in turn are the fibers of  $\tilde{c}$ . Hence, they are the fibers of  $\tilde{c} \circ y$ , which is a fibration in that composition of two fibrations.
- (vii.) This is obvious.

Thus, in the phase space of a superintegrable system  $(M, h, y)$  there are not one but two foliations with leaves which are invariant under the flow of  $X_h$ : the foliation by the invariant tori and a coarser foliation, whose leaves are union of invariant tori – and are thus foliated by them. Such a structure is called a *bifoliation*. We will restrict to the case in which both foliations are fibrations and denote  $M \xrightarrow{y} B \xrightarrow{\tilde{c}} A$  the resulting structure, which is called an (isotropic-coisotropic) *bifibration*. From now on, we will thus abandon our provisional term “integrable  $\mathbb{T}^n$ -fibration” in favor of “bifibration.”

As suggested in Fassò (1996), the bifibrated structure of the phase space of a superintegrable Hamiltonian system can be represented as in Fig. 1: Each fiber of the coisotropic fibration is represented as a flower whose center is the symplectic leaf and whose petals are the invariant tori. The action space  $A$  is the meadow on which the flowers grow. We will use the terms *petals*, *flowers*, and *center of the flowers* as synonyms of,



**Perturbation of Superintegrable Hamiltonian Systems, Fig. 1** The bifibrated structure of the phase space of a superintegrable Hamiltonian system: (a) the three fibrations; (b) pictorial representation: The meadow is the action manifold  $A$ , the flowers are the coisotropic fibers, their centers are the symplectic leaves, and the petals are the invariant tori

respectively, invariant tori, coisotropic leaves, and symplectic leaves. Within each set of generalized action-angle coordinates  $(a, \alpha, p, q)$ , the  $a$ 's are coordinates on the meadow, the  $\alpha$ 's are angles on the petals,  $(\alpha, p, q)$  are local coordinates on the flowers, and  $(p, q)$  are local coordinates on the symplectic leaves.

Our next goal is to enlighten the dynamical meaning of the bifibrated structure. In order to do it, we need to analyze some global issues. On this regard, we note that the doubly foliated structure implies that for a superintegrable system there are different notions of globality: A *local* property pertains to a neighborhood of a point in  $M$ , a *semilocal* property to a neighborhood of a fiber of  $y$  (a group of petals), a *semiglobal* property to a neighborhood of a fiber of  $c$  (a bunch of flowers), and a *global* property to all of  $M$  (all the flowers in the meadow).

**Remark** In the completely integrable case,  $n = d$ ,  $A = B$  and the two fibrations coincide. Thus, complete integrability is a limit case of non-commutative integrability. On the contrary, it is not possible to understand the bifibrated structure of the superintegrable systems starting from that of the completely integrable ones.

### Global Questions

In the hypotheses of Proposition 2, there are various obstructions to the existence of a single, global set of generalized action-angle

coordinates for a bifibration  $M \xrightarrow{y} B \xrightarrow{c} A$ . We refer to (Dazord and Delzant 1987; Duistermaat 1980; Nekhoroshev 1972) for a complete treatment and focus on just two aspects: The (non) existence of an atlas with all transition functions between the local actions and angles equal to the identity and the (non) existence of global coordinates on the symplectic leaves.

The transition functions (7) show that the existence of an atlas with all matrices  $Z = \mathbb{I}$  (the unit matrix) reflects the fact that the fibration  $y$  by the invariant tori is a principal  $\mathbb{T}^n$ -bundle. When this happens, the bifibration  $M \xrightarrow{y} B \xrightarrow{c} A$  is said to have *trivial monodromy*. A bifibration has always trivial monodromy if its action manifold is simply connected. This has the consequence – to be noted in view of perturbation theory – that if a subset  $U$  of  $A$  is simply connected, then  $c^{-1}(U)$  is a principal  $\mathbb{T}^n$ -bundle and has an atlas made of local generalized action-angle coordinates with all matrices  $Z = \mathbb{I}$ . This is true, in particular, for each individual flower.

The presence of the vectors  $z$  in the transition functions (7) is not important for us. Anyway, we note that they can all be chosen  $= 0$  not only if the action space is simply connected, as is obvious, but also if the symplectic structure  $\Omega$  is exact.

The presence of the functions  $\mathcal{F}$  in the transition functions of the angles reflects the presence of a possible obstruction to the topological triviality of the fibration by the invariant tori, that is, the non-existence of a global section. Even though rare among completely integrable systems (Bates 1988), this obstruction is met in many superintegrable systems, actually in an even stronger form: A global section does not exist even for the restriction of the fibration  $y$  to each flower. In such cases, even the single flowers cannot be covered with a single system of coordinates  $(\alpha, p, q)$ . Note that this always happens if the symplectic leaves are compact. In these cases, no semiglobal system of generalized action-angle coordinates can exist – a fact that requires some care in Nekhoroshev theory.

## Dynamics

The doubly fibered structure of a superintegrable Hamiltonian system  $(M, h, y)$  has also a dynamical interpretation.

Obviously, the fibers of the isotropic fibration  $y : M \rightarrow B$  are invariant tori of the system. However, if not all the first integrals have been included among the  $y$ 's, the system could be integrable also with invariant tori of a smaller dimension. For the invariant tori to be considered as uniquely defined by the dynamics, it is necessary that their dimension is the smallest as possible.

## Definition 2

- D4. An integrable Hamiltonian system  $(M, h, y)$  with  $y : M \rightarrow \mathbb{R}^{2d-n}$  is said to have at most  $n$  frequencies.
- D5. An integrable Hamiltonian system  $(M, h, y)$  is said to have  $n$  frequencies if it has at most  $n$  frequencies but not at most  $n - 1$  frequencies.

In order to clarify the dynamical meaning of the fibers of the coisotropic fibration  $c : M \rightarrow A$  (the flowers), and of the symplectic leaves of  $B$ , we need to say something about the frequencies of motions. This will be important for perturbation theory, because resonances depend on frequencies.

A way of defining frequencies is by means of local coordinates. Within a system of generalized action-angle coordinates, the frequencies are  $\omega^{\text{loc}}(a) = \frac{\partial h^{\text{loc}}}{\partial a}(a)$ , see (6). It follows from (7) that, in another system of generalized action-angle coordinates with actions  $a' = Z^{-T}a + z$ , the frequencies of the same motion are  $Z\omega^{\text{loc}}(a)$ . If we want to compare the frequencies of motions on the tori based on two different points  $b, b' \in B$ , we can join these two points by a curve in  $B$ , cover (a lift to  $M$  of) this curve with a number of systems of generalized action-angle coordinates, and “transport” the frequency vector along this chain, multiplying it by the appropriate matrix  $Z$  at each change of chart. If the fibration  $y$  has no monodromy, then the result of this transport depends only on the initial and final points in  $B$ , and not on the chosen curve (and its lift).

Therefore, there is always the possibility of comparing the frequencies of the motions on all the different tori that lie in a subset of the phase space of the type  $c^{-1}(U)$  with  $U$  a connected and simply connected subset of the action space. Note

that this always happens for the tori based on an individual flower. Thus, *all motions inside a flower have the same frequencies*. This conclusion can be made even stronger if, for instance, a Kolmogorov's type condition is satisfied, that is, in each chart, the frequency map is a local diffeomorphism:

$$\det \frac{\partial \omega^{\text{loc}}}{\partial a}(a) \neq 0 \quad \forall a. \quad (9)$$

(This condition is chart-independent because the frequencies transform linearly). In such a case, there is a one-to-one local correspondence between flowers and frequencies. *The flowers coincide, locally, with the set of motions with given frequencies.*

Recall now that the image of a linear motion  $(t, a) \mapsto a + tw \pmod{2\pi}$  on  $\mathbb{T}^n$  is dense in it if and only if its frequency vector  $w \in \mathbb{R}^n$  is not resonant, namely,  $w \cdot v \neq 0$  for all  $v \in \mathbb{Z}^n \setminus \{0\}$ . Given that the irrationals are dense in the reals, if Kolmogorov's condition is satisfied, then for a dense subset  $A_*$  of the action space  $A$ , all invariant tori in the flowers  $c^{-1}(a)$ ,  $a \in A_*$ , are closure of trajectories. In such a situation, the invariant tori  $y^{-1}(b)$ ,  $b \in B$ , are the fibers of the *finest* invariant fibration of  $M$  and the system is certainly integrable with  $n$  frequencies (The distinction between fibration and foliation here is crucial: The finest invariant (possibly singular) foliation of the phase space is given by the orbits). But of course, for this to happen, Kolmogorov's condition is not necessary: Just think of the case in which the frequency map  $\omega^{\text{loc}}(a)$  is constant and nonresonant.

**Remark** *Semilocally*, in a neighborhood of an invariant torus, every superintegrable system can be regarded as a completely integrable one. Consider in fact a set of generalized action-angle coordinates and, restricting its domain if necessary, choose *any* fibration in circles of each coordinate plane  $(q_u, p_u)$ ,  $u = 1, \dots, d - n$ . Then, the  $d$ -dimensional tori constructed as products of the  $n$ -dimensional invariant tori of the system and of  $d - n$  circles in the coordinate planes form a Lagrangian fibration. This operation is not only highly nonunique, because of the arbitrariness in

the choice of the foliations by circles in the coordinate planes, but is in general also only semi-locally defined. Unless the symplectic leaves themselves are fibered by tori of dimension  $d - n$ , there is no fibration by  $d$ -dimensional tori of an entire flower: The construction of these larger tori develops singularities. However, in practice, constructing sets of (semilocally defined)  $d$  commuting integrals of motion and determining the action-angle coordinates of the resulting completely integrable system may be an effective way of finding local sets of generalized action-angle coordinates. For instance, the Delaunay coordinates for Kepler system, which is integrable with one frequency, are exactly of this type.

### Geometric Characterization of Superintegrability

Even though Mischenko and Fomenko's definition of superintegrability is sufficiently general for most cases, it is useful – and clarifying – to adopt a slightly more general, and more geometric, point of view. This will also clarify how Mischenko and Fomenko's assumptions (3) and (4) are related to the setting of Proposition 1.

A *first integral* of a foliation  $\mathcal{F}$  of a manifold  $M$  is any function  $f: M \rightarrow \mathbb{R}$  which is constant on the leaves of  $\mathcal{F}$ , namely,  $L_Y f = 0$  for all vector fields  $Y$  tangent to the leaves of  $\mathcal{F}$ . If  $\mathcal{F}$  is a fibration  $y: M \rightarrow B$ , then its first integrals are the lifts to  $M$  of functions defined on the base  $B$ .

Consider now a foliation  $\mathcal{F}$  of a symplectic manifold  $(M, \Omega)$ . First, note that  $f: M \rightarrow \mathbb{R}$  is a *first integral* of  $\mathcal{F}$  if and only if  $X_f$  is everywhere *symplectically orthogonal* to  $\mathcal{F}$ , namely,  $X_f(x) \in (T_x \mathcal{F}_x)^\perp$  for all  $x \in M$  (this follows from the identity  $L_Y f = \Omega(Y, X_f)$ ).

Next, the *polar* of  $\mathcal{F}$ , if it exists, is defined as the unique foliation  $\mathcal{F}^\perp$  of  $M$  with the property that the tangent spaces of its leaves are the symplectic orthogonals of the tangent spaces of the leaves of  $\mathcal{F}$ . If the leaves of  $\mathcal{F}$  have dimension  $n$ , then the leaves of  $\mathcal{F}^\perp$  have dimension  $\dim M - n$ . If  $\mathcal{F}$  is isotropic (By saying that a foliation or a fibration is (co)isotropic, we mean that its leaves or fibers are (co)isotropic), then  $\mathcal{F}^\perp$  is coisotropic; hence, each leaf of  $\mathcal{F}^\perp$  is a union of leaves of  $\mathcal{F}$ . A foliation  $\mathcal{F}$  which has a polar is called *symplectically complete*



(Dazord and Delzant 1987). The terms *bifoliation*, *bifibration* (if both  $\mathcal{F}$  and  $\mathcal{F}^\perp$  are fibrations), and under some stronger condition, *dual pair* are also used to denote the pair  $(\mathcal{F}, \mathcal{F}^\perp)$ . For a general treatment of these objects and the proof of (statements equivalent to) Proposition 6 below, see Dazord and Delzant (1987), Sect. 9 of Chap. 3 of Libermann and Marle (1987), and Sect. 11.3 of Ortega and Ratiu (2004).

The following are examples of symplectically complete foliations:

- E1. Any Lagrangian foliation, which coincides with its own polar.
- E2. The orbits of a Hamiltonian vector field  $X_h$ : The polar is given by the (connected components of the) level sets of  $h$  (If a vector field  $Y$  is tangent to the level sets of  $h$ , then  $\Omega(Y, X_h) = L_Y h = 0$ ).
- E3. The orbits of a Hamiltonian action and the level sets of its momentum map are polar to each other; see (Libermann and Marle 1987; Marsden and Ratiu 1994).

However, not all isotropic foliations have a polar (Woodhouse 1994). We limit ourselves to state some facts:

**Proposition 6** *Let  $\mathcal{F}$  be a foliation of a symplectic manifold  $(M, \Omega)$ .*

- B1.  *$\mathcal{F}$  is symplectically complete if and only if the Poisson brackets of any two first integrals of  $\mathcal{F}$  is a first integral of  $\mathcal{F}$ .*

*Consider now a foliation  $\mathcal{F}$  of  $M$  which is a fibration  $\gamma : M \rightarrow B$ .*

- B2.  *$\mathcal{F}$  is symplectically complete if and only if there exists a Poisson structure on  $B$  such that  $\gamma : M \rightarrow B$  is a Poisson morphism.*
- B3. *Assume  $\mathcal{F}$  is symplectically complete. Then its leaves are isotropic if and only if the Poisson structure of  $B$  has rank =  $2 \dim B - \dim M$ .*
- B4. *If the leaves of  $\mathcal{F}$  are isotropic, then they are generated by the Hamiltonian vector*

*fields of the lifts to  $M$  of the (local) Casimirs of  $B$ .*

Thus, in the definition of noncommutative integrability, condition (3) means that the fibration by the invariant tori has a polar foliation, and condition (4) that the invariant tori are isotropic. Note that property B3 implies that the maximal number of independent (local) Casimirs of  $B$  is  $n$ . Lastly, property B4 explains that, if conditions (3) and (4) are satisfied, then (semiglobal) sets of commuting dynamical symmetries (the vector fields  $Y_1, \dots, Y_n$  of Proposition 1) are provided by the Hamiltonian vector fields of the lifts to  $M$  of any set of  $n$  independent (local) Casimirs of  $B$ .

This suggests a geometric generalization of Mischenko and Fomenko's noncommutative integrability: requiring the existence of a symplectically complete fibration  $\bar{\gamma} : M \rightarrow \bar{B}$  by isotropic tori, with  $\bar{B}$  any  $(2d - n)$ -dimensional (Poisson) manifold instead of an open subset of  $\mathbb{R}^{2d - n}$ . Locally, of course, in sets  $\bar{\gamma}^{-1}(\bar{U})$  with  $\bar{U} \subset \bar{B}$  diffeomorphic to an open set in  $\mathbb{R}^{2d - n}$ , this fibration can be described as in Mischenko and Fomenko's characterization, but these local descriptions may hinder the global structure of the bifibration.

## Examples

We describe now a few examples of superintegrable systems and their bifibrated structure.

1. *Maximally superintegrable systems.* Assume that a Hamiltonian vector field  $X_h$  on a symplectic manifold  $M$  has all its orbits periodic. The orbits of  $X_h$  are diffeomorphic to  $S^1$  and, being one-dimensional, are isotropic. Example E3 above implies that the orbits are the leaves of a symplectically complete foliation and that the leaves of the polar foliation are (the connected components of) the  $2d - 1$ -dimensional level sets of  $h$ . It is known that if, for instance, the period of the motion is a continuous function of the initial datum, see Fassò et al. (2005), then the orbits of  $X_h$  are the fibers of a fibration  $\hat{\gamma} : M \rightarrow \hat{B}$ , with base a



manifold of dimension  $2d - 1$ . (However,  $B$  might not be an open subset of  $\mathbb{R}^{2d-1}$ , and we are, therefore, within the “geometric” characterization of superintegrability seen at the end of the previous section. We limit ourselves to mention that two examples of this situation are offered by the three-dimensional isotropic oscillator Karasev and Maslov (1993) and the spatial Kepler system.)

The Hamiltonian is a choice of the (possibly local) only Casimir. The flowers are the connected components of the level sets of  $h$ , and each of them has one-dimensional petals and a  $(2d - 2)$ -dimensional center. The action manifold is one-dimensional and can be identified with the set of values of  $h$ , if  $h$  has connected fibers. The structure and interpretation of the symplectic leaves depend on the system. The next example details the Kepler system. A detailed treatment of another example, the resonant 1:1 oscillator or planar isotropic oscillator, can be found in Fassò (2005); Karasev and Maslov (1993). For the  $m:n$  resonant oscillators, see Holm and Vizman (2012).

2. *The planar Kepler system.* First consider the planar Kepler system. The Hamiltonian is  $h(q, p) = \frac{\|p\|^2}{2} - \frac{1}{\|q\|}$  in  $(\mathbb{R}^2 \setminus \{0\}) \times \mathbb{R}^2 \ni (q, p)$ . The flow is periodic in the subset  $M = M_+ \cup M_-$  of the phase space where  $H < 0$  and the angular momentum  $G := q_1 p_2 - q_2 p_1$  is nonzero, positive in  $M_+$ , and negative in  $M_-$ . Motions in  $M$  take place along the anti-clockwise ( $G > 0$ ) and clockwise ( $G < 0$ ) Keplerian ellipses, which have major semiaxis  $L = \frac{1}{\sqrt{-2H}}$  and eccentricity  $\sqrt{1 - \frac{G^2}{L^2}} \in [0, 1)$ . The functions  $h$ ,  $L$  and  $G$  are integrals of motion and such are also the two components  $A_1 = p_2 G - \frac{q_1}{\|q\|}$  and  $A_2 = -p_1 G - \frac{q_2}{\|q\|}$  of the Laplace vector  $A$  ( $A$  points to the pericenter of the ellipse, and  $\|A\|$  equals the eccentricity). The fibration by the periodic orbits is given by the map  $y : M \rightarrow \mathbb{R}^3$  with components  $y_1 = L A_1$ ,  $y_2 = L A_2$ ,  $y_3 = G$ . Precisely,  $y : M \rightarrow B = \{\tilde{y} \in \mathbb{R}^3 : \tilde{y}_3 \neq 0\}$  is a surjective submersion and has connected fibers ( $y_1$  and  $y_2$  fix semiaxis and orientation of the Keplerian

ellipse and  $y_3$  its eccentricity and orientation). Since the Poisson brackets of the components of  $y$  are  $\{y_1, y_2\} = y_3$  and cyclic permutations, the matrix  $P(\tilde{y})$  is the antisymmetric matrix of the vector product with  $\tilde{y}$  and has rank  $2 = 3 - 1$  for all  $\tilde{y} \neq 0$ , hence in all of  $B$ . Since  $\|y\|^2 = L^2 \|A\|^2 + G^2 = L^2$  is a function of  $h$ , a choice of the Casimir of  $B$  is the Euclidean norm and the symplectic leaves are the sets  $\|\tilde{y}\| = \text{const}$  in  $B$ , namely, the union of the two half spheres  $S_{\pm} = \{\tilde{y} \in \mathbb{R}^3 : \|\tilde{y}\| = \text{const} > 0, \tilde{y}_3 \gtrless 0\}$  (at their common boundary – the equator of the sphere – there are the collision orbits: The noncompactness of the symplectic leaves is due to the fact that motions at zero angular momentum are not periodic). Note that the circular orbits sit on the two poles  $\tilde{y}_3 = \pm \|\tilde{y}\|$  of the symplectic leaves and are regular fibers of the fibration by the periodic orbits.

Local generalized action-angle coordinates for the planar Kepler system are provided by the well-known Delaunay elements  $(L, G, \ell, g)$  (Arnold et al. 2006), where  $\ell$  is the mean anomaly measured from the pericenter and  $g$  is the argument of the pericenter, which are defined in the subset of  $M$  where  $|G| < L$ , namely not on the circular orbits (the angles  $g$  and  $\ell$  are not individually defined on the circular orbits, even though their sum (resp. difference) is defined on those with  $G > 0$  (resp.  $G < 0$ )).  $L$  is the action,  $\ell$  is the angle, and  $(G, g)$  are coordinates on the symplectic leaves with the two poles  $\tilde{y}_3 = \pm \|\tilde{y}\|$  removed. The singularity at the poles on the symplectic leaves is due to the fact that  $(G, g)$  are cylindric coordinates. A global set of generalized action-angle coordinates for each set  $M_{\pm}$  is provided by the Poincaré' elements  $(L, \lambda, \eta, \xi)$ , where  $\lambda = g \pm \ell$ ,  $\eta = \frac{\sqrt{2} y_1}{\sqrt{\|\tilde{y}\| \pm y_3}}$ ,  $\xi = -\frac{\sqrt{2} y_2}{\sqrt{\|\tilde{y}\| \pm y_3}}$ .

3. *The spatial Kepler system.* The structure of the spatial Kepler system is similar. The flow is periodic in the subset  $M$  of the phase space  $(\mathbb{R}^3 \setminus \{0\}) \times \mathbb{R}^3 \ni (q, p)$  where the Hamiltonian  $h = \frac{\|p\|^2}{2} - \frac{1}{\|q\|}$  is negative and the angular momentum  $m = q \times p$  is nonzero. The base

$B$  is five-dimensional; here too, the Hamiltonian or the greater semiaxis  $L = \frac{1}{\sqrt{-2h}}$  can be taken as the only Casimir, and the four-dimensional symplectic leaves are diffeomorphic to the Cartesian product of a two-dimensional sphere and of a two-dimensional sphere minus a point,  $S^2 \times (S^2 \setminus \{p\})$  with  $p \in S^2$ . (The excluded sphere  $S^2 \times \{p\}$  corresponds to the orbits of energy  $h$  and zero angular momentum: Again, the symplectic leaves are not compact due to collisions with the force center.) This picture can be deduced, for instance, from the treatment in Section II.3 of Cushman and Bates (2015), even though the approach there is somewhat different in that it does not refer to non-commutative integrability (We also note that the sets that here are interpreted as symplectic leaves are there described as diffeomorphic to  $\{(u, v) \in \mathbb{R}^3 \times \mathbb{R}^3 : \|u\| = \|v\| = L, v \neq -u\}$ ). The points in the symplectic leaves fix the placing in space of the plane of the keplerian ellipse, its orientation in that plane, and its eccentricity.

4. *The point in a central force field.* The phase space is  $(\mathbb{R}^3 \setminus \{0\}) \times \mathbb{R}^3 \ni (q, p)$  (or  $\mathbb{R}^3 \times \mathbb{R}^3$  if the potential energy  $V$  is regular at the origin), and assuming that the point has unit mass, the Hamiltonian is  $h(p, q) = \frac{1}{2}\|p\|^2 + V(\|q\|)$ . The system has four integrals, the energy  $h$  and the three components  $(m_x, m_y, m_z)$  of the angular momentum vector  $m = q \times p$ . The map  $y = (h, m)$  is a submersion except where  $q$  is parallel to  $p$ , namely to  $\dot{q}$ , and where  $V'(\|q\|) = 0, p = 0$ . Assume that in an open set  $M$  where it is a submersion,  $(h, m)$  has compact and connected fibers (this depends on the potential energy  $V$ ). Then,  $(h, m)$  satisfies the hypotheses of Proposition 1 with  $d = 3$  and  $n = 2$  because

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & m_z & -m_y \\ 0 & -m_z & 0 & m_x \\ 0 & m_y & -m_x & 0 \end{pmatrix} \quad (10)$$

has rank two wherever  $m \neq 0$ . With a minor abuse of notation, (possibly local) Casimirs of  $B$  are the Hamiltonian  $h$  and the function  $\|m\|$ .

The symplectic leaves of  $B$  are the two-dimensional spheres  $h = \text{const}, \|m\| = \text{const}$ ; for fixed energy, they can be identified with the spheres of constant norm of the angular momentum vector. The flowers are diffeomorphic to  $\text{SO}(3) \times S^1$ , and the fibration of a flower over a symplectic leaf can be viewed as a fibration  $\text{SO}(3) \times S^1 \rightarrow S^2$  which involves the Hopf fibration  $\text{SO}(3) \rightarrow S^2$  and is nontrivial (see Proposition 4 in Bates and Fassò (2016)). Local sets of generalized action-angle coordinates are explicitly known only for special choices of the potential, for instance, for the so-called Henon's isochronous potential  $V(r) = \frac{1}{1+\sqrt{1+r^2}}$ , see (Ramond and Perez 2021); however,  $\|m\|$  can always be taken as an action.

5. *The Euler top.* The Euler top (or Euler-Poinsot system) is the rigid body with a fixed point and no external torques acting on it. The phase space can be identified with  $\text{SO}(3) \times \mathbb{R}^3 \ni (R, m^b)$ , with the matrix  $R$  that fixes the orientation of the body and  $m^b = (m_1, m_2, m_3)$  being the body-angular momentum vector. The system has four independent integrals of motion: the kinetic energy  $h(m^b) = \frac{m_1^2}{2I_1} + \frac{m_2^2}{2I_2} + \frac{m_3^2}{2I_3}$ , where  $I_1, I_2$ , and  $I_3$  are the principal moments of inertia, and the three components of the spatial angular momentum vector  $m^s := Rm^b$ . Assume  $I_1 < I_2 < I_3$  and consider the energy-momentum map  $y = (h, m^s)$ . The matrix of the Poisson brackets of the components of  $y$  has the form (10), and  $y$  satisfies (3) and (4) at all points where  $m^b \neq 0$ . Let  $M$  be the complement in the phase space of the set formed by the equilibria ( $m^b = 0$ ), by the stationary rotations about the principal axes of inertia ( $m^b$  parallel to one of the basis vectors of  $\mathbb{R}^3$ ) and by the stable and unstable manifolds of the stationary rotations relative to the middle axis of inertia. Then,  $y : M \rightarrow \mathbb{R}^4$  is a submersion and, some technicalities aside ( $M$  has two connected components, and  $y : M \rightarrow \mathbb{R}^4$  is a submersion with fibers diffeomorphic to  $\mathbb{T}^2 \cup \mathbb{T}^2$ , hence not connected),  $M$  is the union of four open subsets  $M_i, i = 1, \dots, 4$ , such that  $y|_{M_i} : M_i \rightarrow B_i \subset \mathbb{R}^4$  is an integrable  $\mathbb{T}^2$ -fibration.

Two independent (local) Casimirs for each  $y|_{M_i}$  are  $h$  and  $\|m^s\|$ , and the symplectic leaves  $h = \text{const}$ ,  $\|m^s\| = \text{const}$  in each  $B_i$  are two-dimensional spheres; for given  $h$ , they can be identified with the two-dimensional spheres of constant norm of the angular momentum vector in space. The flowers are the fibers of the submersions  $(h, \|m^s\|) : M_i \rightarrow A_i \subset \mathbb{R}^2$  and are diffeomorphic to  $\text{SO}(3) \times S^1$  (the matrix  $R$  is unrestricted while  $m^b \in \mathbb{R}^3$  belongs to one of the two circles resulting from the intersection of a sphere and of an ellipsoid: This very famous picture appears virtually in every textbook on Mechanics). Each  $A_i$  is contractible and there is no monodromy, but here too the fibration  $\text{SO}(3) \times S^1 \rightarrow S^2$  by the invariant tori of each individual flower involves a Hopf fibration and is nontrivial. Local systems of generalized action-angle coordinates can be constructed out of the so-called Andoyer-Deprit coordinates, which are local action-angle coordinates of the fibrations by the three-dimensional tori given by the three commuting integrals  $m_3$ ,  $\|m^s\|$  and one of the components of  $m^s$ . We refer to (Kozlov 1974) for a study of their properties and mention that one of the actions is  $\|m^s\|$ , the other is a function of  $h$  and  $\|m^s\|$ , and the local coordinates  $(p, q)$  can be taken as symplectic cylindric coordinates on the spheres  $\|m^s\| = \text{const}$ . For further details and references, see (Fassò 1996).

6. *Symmetry and superintegrability.* Superintegrability is often linked to the existence of a “large” symmetry group, the fibration by the invariant tori being given either by the momentum map or by the energy-momentum map (we assume some previous knowledge of this topic; see (Libermann and Marle 1987; Marsden and Ratiu 1994)).

Assume there is a Hamiltonian action of a Lie group  $G$  on a symplectic manifold  $M$ , with coadjoint equivariant momentum map  $J : M \rightarrow \mathfrak{g}^*$ . As noticed in example E3 above, the level sets of  $J$  form a symplectically complete foliation, with polar given by the group orbits. Therefore, if  $J : M \rightarrow J(M)$  is a fibration with compact and connected fibers, what remains to be checked in order to conclude that its level sets are the invariant tori of a

superintegrable system is their isotropy. Since  $J$  is a Poisson morphism from the symplectic manifold to  $\mathfrak{g}^*$ , statement B3 of Proposition 6 implies that the isotropy of its fibers depends on the dimension of the coadjoint orbits  $O_\mu$ ; specifically, they are isotropic if and only if

$$\dim O_\mu = 2d - 2n \quad \forall \mu \in J(M). \quad (11)$$

In this situation, the flowers are the group orbits, the symplectic leaves of the base  $J(M)$  can be identified with the (connected components of the) coadjoint orbits, and the petals are the orbits of a maximal torus of  $G$ . An example of this situation is offered by the dynamically symmetric Euler top, with moments of inertia  $I_1 = I_2 \neq I_3$ , which has symmetry group  $\text{SO}(3) \times S^1$  (rotations in space and around the symmetry axis). The Hamiltonian equals  $\frac{\|m\|^2}{2I_1} + \frac{(I_1 - I_3)m_3^2}{2I_1 I_3}$ , and the fibration by the invariant two-dimensional tori can be described via the momentum map  $(m_3, m^s)$  of this action instead of the (energy-momentum) map  $(h, m^b)$  considered in Example 4.

If the Hamiltonian  $h$  is independent of the momentum map  $J$ , one should consider the energy-momentum map  $(h, J)$ , which defines an invariant foliation which is finer than  $J$ . This case can be reduced to the previous one by observing that  $(h, J)$  is the momentum map of the action of the direct product  $\mathbb{R} \times G$  given by  $(t, g).x = \Phi_t^{X_h}(g.x)$ . Thus, if the energy momentum map is a fibration with compact connected fibers, then the isotropy condition becomes  $\dim O_\mu = 2 + 2 \dim J(M) - \dim M$ .

**Remark** If the action of  $G$  is free, then the  $G$ -orbits, namely the flowers, are diffeomorphic to  $G$  and  $\dim \mathfrak{g} = \dim G = 2d - n$ . Define the *rank* of the Lie algebra  $\mathfrak{g}$  of  $G$  as the codimension of its (generic) coadjoint orbits. In the current hypotheses,  $\text{rank } \mathfrak{g} = \dim \mathfrak{g} - \dim O_\mu = (2d - n) - (2d - 2n) = n$  and (11) can be written as

$$\dim M = \dim \mathfrak{g} + \text{rank } \mathfrak{g}.$$

This, not (10), is the condition used to define noncommutative integrability in Mischenko and Fomenko (1978), with  $\mathfrak{g}$  the  $(2d - n)$ -dimensional

noncommutative Lie algebra formed by the first integrals. See also Sect. 3.7 of Arnold and Givental (2001).

## Perturbations of Superintegrable Systems

### Introduction

We consider now a small perturbation of a superintegrable system with bifibration  $M \xrightarrow{y} B \xrightarrow{\tilde{c}} A$  and, employing our usual notation, Hamiltonian  $k = \hat{k} \circ c$  with a function  $\hat{k} : A \rightarrow \mathbb{R}$ . The Hamiltonian of the perturbed system is

$$h = k + \varepsilon f \quad (12)$$

with  $f : M \rightarrow \mathbb{R}$  and  $\varepsilon \geq 0$  a small parameter. The goal is to understand the long-term properties of the dynamics of the perturbed system for small  $\varepsilon$ .

Systems of this type emerge not only as perturbations of a given superintegrable system, but also, as is common in Celestial Mechanics, as approximations of nonintegrable systems. For instance, the so-called  $n + 1$  planetary system is formed by a star of mass  $m_S$  and by  $n \geq 2$  planets of masses  $\ll m_S$  and can be viewed (after quotienting away the three translational degrees of freedom, or equivalently working in a star-centric frame) as a small perturbation of the system formed by  $n$  noninteracting systems star-planet, namely,  $n$  noninteracting Kepler systems. In the spatial case, this is a superintegrable system with  $d = 3n$ , hence  $n$ -dimensional invariant tori and  $4n$ -dimensional symplectic leaves (diffeomorphic to the product of  $n$  copies of  $S^2 \times (S^2 \text{ minus a point})$ , see Example 2 in the previous section), while in the planar case  $d = 2n$  and the symplectic leaves have dimension  $2n$  (now  $n$  copies of a sphere minus the equator).

The development of perturbation theory in the nineteenth century was largely motivated by the study of perturbations of the planetary system – hence of a superintegrable system – and versions of both KAM and Nekhoroshev theorems for perturbations of superintegrable systems appeared in the early stages of the development of these theories. These analyses were performed by considering the system in the domain of a set of

generalized action-angle coordinates, where Hamiltonian (12) has the form

$$h^{\text{loc}}(a, \alpha, p, q) = k^{\text{loc}}(a) + \varepsilon f^{\text{loc}}(a, \alpha, p, q) \quad (13)$$

and Hamilton equations are

$$\begin{aligned} \dot{a} &= -\varepsilon \frac{\partial f^{\text{loc}}}{\partial \alpha}, & \dot{\alpha} &= \omega^{\text{loc}} + \varepsilon \frac{\partial f^{\text{loc}}}{\partial a}, \\ \dot{p} &= -\varepsilon \frac{\partial f^{\text{loc}}}{\partial q}, & \dot{q} &= \varepsilon \frac{\partial f^{\text{loc}}}{\partial p} \end{aligned}$$

with  $\omega^{\text{loc}} = \frac{\partial k^{\text{loc}}}{\partial a}$ . In perturbation theory, systems of this type, with  $d > n$ , are often referred to as perturbations of “properly degenerate” systems. The reason is that, if the  $q_j$  are chosen to be angles (see the Remark at the end of the subsection *Dynamics*), then the coordinates  $(a, \alpha, p, q)$  can be seen as a system of action-angle coordinates of a completely integrable system whose Hamiltonian  $k^{\text{loc}}$  depends only on  $n$  of the  $d$  actions, and Kolmogorov’s nondegeneracy condition is violated everywhere. Clearly, the motion of the coordinates  $(p, q)$ , namely the motion of the system along the symplectic leaves, takes place with speed order  $\varepsilon$ . Nevertheless, the times of interest in perturbation theory are very long compared to  $\frac{1}{\varepsilon}$ , and the slowness of this motion, per se, does not imply any confinement. In fact, the main difference with respect to the completely integrable case is that perturbations of superintegrable systems may produce a variety of behaviors of the motion along the symplectic leaves, which may be either regular or chaotic and is not yet completely understood. A description of existing results can be found in Chap. 6 of Arnold et al. (2006) and in the reviews (Guzzo 2007; Hanßmann 2015).

Even though historically KAM theory predates Nekhoroshev theory, we describe first the latter, which gives a more comprehensive description of the dynamics.

### Semiglobal Approach

Preliminarily, we make some remarks related to the possibility that the superintegrable system does not have a single, global system of generalized action-angle coordinates.

First we note that, since any KAM or Nekhoroshev result aims at showing that the

actions – namely the projection (This projection is well defined because  $c \circ \gamma : M \rightarrow A$  is a fibration) of motions onto the action space – remain nearly constant for the time-scale of interest, the analysis can always be restricted to subsets of the phase space that project over very small sets in the action space, which in particular can be taken simply connected and diffeomorphic to open subsets of  $\mathbb{R}^n$ . Therefore, monodromy of the superintegrable system, if present, can be ignored in both KAM and Nekhoroshev studies (This might not be true if Arnold diffusion is concerned, because the system might move, e.g., around “holes” of the action space due to singularities of the system. However, we are not aware of any result in this direction). It is thus sufficient to cover the action space  $A$  with simply connected open sets  $A_*$  diffeomorphic to open sets in  $\mathbb{R}^n$  and give “semiglobal” statements of KAM and Nekhoroshev theorems in the sets  $M_* = c^{-1}(A_*)$ . For simplicity, we may even identify each  $A_*$  with an open subset of  $\mathbb{R}^n$ . However, it is *not* possible in general to assume that the subsets  $M_* := c^{-1}(A_*)$  have a single, global system of generalized action-angle coordinates (just think of the case of compact symplectic leaves). As we shall see, this is not an issue for KAM theory, where normal forms built in the domain of some system of generalized action-angle coordinates (hence “semilocally”) appear to be sufficient, but things are different for Nekhoroshev theory.

Now, Nekhoroshev theory is based on the construction of normal forms adapted to the local resonance properties of the unperturbed system. As we have seen, for a superintegrable system the frequencies of motions are a property not of the individual isotropic tori (the petals) but of the coisotropic flowers. Therefore, the natural sets where normal forms adapted to the resonance properties of the superintegrable system should be constructed are open neighborhoods of the flowers. The possibility of constructing “semiglobal” normal form, proved in Fassò (1995), is reviewed in the Appendix.

Furthermore, the validity of Nekhoroshev theorem requires the “real analyticity” of the Hamiltonian  $h : M \rightarrow \mathbb{R}$ . To give a meaning to it, first require that the bifibration  $M \xrightarrow{\gamma} B \xrightarrow{\tilde{c}} A$  is real analytic (in the sense that the transition functions

between its local systems of generalized action-angle coordinates are all real analytic) and then define as “real analytic” any function  $f : M \rightarrow \mathbb{R}$  whose all local representatives in the charts of such an atlas are real analytic. (It is sufficient to check both conditions for a single atlas with generalized action-angle coordinates, not for all such atlases, because by the Grauert-Morrey theorem any smooth manifold has a unique real analytic structure compatible with its differentiable structure).

Thus, a “semiglobal” statement of Nekhoroshev theorem can be given for real analytic perturbations of a superintegrable system on a real analytic bifibration  $M_* \xrightarrow{\gamma} B_* \xrightarrow{\tilde{c}} A_*$  with  $A_*$  an open, simply connected subset of  $\mathbb{R}^n$  (but  $B_*$  and  $M_*$  manifolds).  $M_*$  possesses an atlas with generalized action-angle coordinates with all transition matrices between the actions equal to the identity, and we will (tacitly) use such an atlas. In order to simplify things a little bit, we will also assume that all chart domains of this atlas project over the entire open set  $A_* \subseteq \mathbb{R}^n$ , so that the action coordinates of each and every chart of  $M_*$  provide a global coordinate system for  $A_*$  and we need not distinguish between them. In this situation, the functions  $k$  and  $\hat{k}$  can be identified with their expression in any of the considered coordinate systems, and it is possible to define the frequency vector as the map  $\omega : A_* \rightarrow \mathbb{R}^n$  such that

$$\omega(a) = \frac{\partial k}{\partial a}(a).$$

Hence, nondegeneracy, convexity, etc. are defined as in local coordinates. Note that, in this way, we may also introduce a distance  $d$  in  $A_*$  (for instance, the Euclidean one in  $\mathbb{R}^n$ ).

### Nekhoroshev Theorem for Superintegrable Systems

The original statement and proof that Nekhoroshev gave of his own celebrated theorem was for a class of systems that includes (a semilocal formulation of) superintegrable systems. Specifically, in Nekhoroshev (1977) Nekhoroshev considered systems that he called “systems with parameters,” which in symplectic coordinates  $(a, \alpha, p, q) \in \mathbb{R}^n \times \mathbb{T}^n \times \mathbb{R}^m \times \mathbb{R}^m$ ,



with symplectic form  $\sum_{i=1}^n da_i \wedge dx_i + \sum_{u=1}^m dp_u \wedge dq_u$  ( $n \geq 1, m \geq 0$ ), have Hamiltonian of the form (13). The “main theorem” in Sect. 4.4 of Nekhoroshev (1977) states that if  $k^{\text{loc}}$  and  $f^{\text{loc}}$  are real analytic and if  $k^{\text{loc}}$  is convex (or more generally quasi-convex or steep), then the actions  $a$  remain nearly constant for times which are the minimum between the exponentially long Nekhoroshev timescale

$$T_{\text{exp}} \sim \exp(\varepsilon^{-\beta})$$

( $\beta = \frac{1}{2n}$  if  $k^{\text{loc}}$  is either convex or quasiconvex) and the escape time of the coordinates  $(p, q)$  of the motion from the coordinate domain (More precisely: For any pair of open, relatively compact subsets  $U$  of the domain of the actions  $a$  and  $C$  of the domain of the coordinates  $(p, q)$ , there exist positive constants  $\delta$  and  $\varepsilon_* > 0$  (which depend only on  $k, f, U$ , and  $C$ ) such that if  $\varepsilon < \varepsilon_*$ , then any motion  $t \mapsto z_t = (a_t, p_t, q_t)$  of the system with initial datum  $z_0 \in U \times \mathbb{T}^n \times C$  and with  $a_0$  at distance  $\geq \delta$  from  $\partial A_*$ , satisfies  $\|a_t - a_0\| \leq \delta \varepsilon^\beta$  for  $|t| \leq \min(T_{\text{exp}}, T_{z_0}^{\text{chart}})$  where  $T_{z_0}^{\text{chart}}$  is the escape time of the coordinates  $(p, q)$  from  $C$ . The restriction to a relatively compact (namely, with compact closure) subset of the coordinate domain may be necessary because the Hamiltonian and its derivatives might diverge at the boundary).

The limitation of this result, when applied to a superintegrable system, is the presence of the escape time from the chart domain. Given that the coordinates  $(p, q)$  move with speed of order  $\varepsilon$ , they might in principle leave the considered chart domain within a time of order  $\frac{1}{\varepsilon}$ , too short compared to the Nekhoroshev timescale. In practice, as will be discussed below, it might be difficult or even impossible to control the motion of the coordinates  $(p, q)$ , which might be chaotic, and in such a case there is no knowledge at all of the escape time from the chart domain, besides the a priori estimate  $\sim \frac{1}{\varepsilon}$ . In such a situation, the estimate on the variation of the actions provided by the theorem is granted only for a timescale  $\frac{1}{\varepsilon}$ .

However, building semiglobally (rather than semilocally) the normal forms that enter the proof of Nekhoroshev theorem, it is possible to

prove that the previous estimate is valid with the escape time from the chart replaced by the escape time from the symplectic leaves; see (Fassò 1995).

**Proposition 7** (Nekhoroshev theorem for superintegrable systems) *Consider a superintegrable system with real analytic and convex (or quasi-convex) Hamiltonian  $\hat{k}: A \rightarrow \mathbb{R}$  and bifibration  $M \xrightarrow{y} B \xrightarrow{\bar{c}} A$ . For any real analytic function  $f: M \rightarrow \mathbb{R}$  and any pair of open and relatively compact sets  $A_* \subseteq A$  and  $K \subseteq B$  such that  $\tilde{c}(K) \subseteq A_*$ , there exist positive constants  $\varepsilon_*, \delta, t_*$  such that if  $|\varepsilon| \leq \varepsilon_*$ , then all motions  $t \mapsto z_t$  of the system of Hamiltonian  $\hat{k} \circ c + \varepsilon f$  with initial datum  $z_0 \in y^{-1}(K)$ , and such that  $c(z_0)$  is at distance  $> \delta$  from  $\partial A_*$ , satisfy*

$$d(c(z_t), c(z_0)) < \delta \left( \frac{\varepsilon}{\varepsilon_*} \right)^\beta \text{ for time } |t| < \min \left( T_{\text{exp}}, T_{z_0}^{\text{SL}} \right) \quad (14)$$

with  $T_{\text{exp}} = t_* \exp \left( \left( \frac{\varepsilon_*}{\varepsilon} \right)^\beta \right)$ ,  $\beta = \frac{1}{2n}$ , and  $T_{z_0}^{\text{SL}}$  the escape time of the motion from  $y^{-1}(K)$ .

There are two cases:

1. If the symplectic leaves of  $B$  are compact, then it is possible to take the compact set  $K$  in such a way that it is a union of symplectic leaves. Since no escape is possible along the symplectic leaves, the escape from  $K$ , if any, can only be due to the movement of the actions  $a$ . Thus,  $T_{z_0}^{\text{SL}} > T_{\text{exp}}$  and Proposition 7 provides stability of the actions  $a$  for the entire Nekhoroshev timescale  $T_{\text{exp}}$ .
2. The noncompactness of the symplectic leaves may be due either to the presence of singularities (e.g., the equator of collisions in Kepler system) or the fact that the symplectic leaves “go to infinity.” Excluding that the perturbed system falls into a singularity or runs to infinity within the Nekhoroshev time  $T_{\text{exp}}$  is a dynamical question which depends on the perturbation and, in general, can at best be guaranteed only for initial conditions in some subset of the



phase space. For initial data in such a subset, again Proposition 7 provides stability of the actions  $a$  for the timescale  $T_{\text{exp}}$ .

Thus, in both cases the Nekhoroshev theorem for superintegrable systems ensures that, on the time  $T_{\text{exp}}$ , the system remains very near the flower it started from. What remains to be understood are the properties of the dynamics in the flowers, in particular along the symplectic leaves.

### Remarks

- (i) If the study of the motions along the symplectic leaves allows to conclude that the system remains inside a chart domain, then there is no need to resort to the semiglobal construction of Proposition 7 and Nekhoroshev's original semilocal formulation can be used. Some cases are discussed in the next section.
- (ii) Proposition 7 was proved in Fassò (1995) by showing that the original scheme of proof developed by Nekhoroshev in Nekhoroshev (1977) can be adapted using semiglobal – instead of semilocal – normal forms. Nekhoroshev's original proof is based on the decomposition of the phase space in regions characterized by certain resonance properties and on the construction of normal forms adapted to them (something about this construction is said in the next section). A different proof, but only for the case in which the flowers are the orbits of a compact group action (and hence are in particular compact), is given in Blaom (2001) using a version of the proof of Nekhoroshev theorem, due to Lochak Lochak (1992), which requires only the construction of maximally resonant normal forms. This approach makes the proof somewhat easier (see the Appendix) but has the important limitation that – avoiding the construction of all other resonant and nonresonant normal forms – cannot provide a detailed description of the motions along the symplectic leaves.

### Motions Along the Symplectic Leaves

The main dynamical question left open from the Nekhoroshev theorem for superintegrable

systems concerns the properties, on the timescale  $T_{\text{exp}}$ , of the motion along the symplectic leaves, beginning with whether these motions are regular or chaotic. These motions can be investigated using the normal forms constructed within the proof of Nekhoroshev theorem and as it turns out, their properties are crucially affected by the resonance properties of motions. (In this sense, in the author's opinion, Nekhoroshev theorem is not only a statement about the stability of action, but also a theory of resonant motions of nearly integrable Hamiltonian systems). There is not yet a complete understanding of this topic except, to a certain extent, when  $d = 3$  and  $n = 2$ .

To describe the known results, we must recall a few details of the proof of Nekhoroshev theorem; in so doing, we refer to the notation of Proposition 7. Within the proof, the action space  $A_*$  is subdivided in regions characterized by (approximate) resonance properties. Due to the exponential decay of the Fourier components of analytic functions, only resonances of order not greater than a “cutoff”

$$N(\varepsilon) \sim \left(\frac{\varepsilon_*}{\varepsilon}\right)^\beta$$

affect the dynamics on times up to  $T_{\text{exp}} \sim \exp(N(\varepsilon))$  and are considered; higher order resonant motions are indistinguishable, on that timescale, from non-resonant motions. The proof considers all lattices  $L$  of  $\mathbb{Z}^n$  which have a set of generators of  $L^1$ -norm  $\leq N(\varepsilon)$  and, for each of them, a small neighborhood  $A_L \subset A_*$  of the  $L$ -resonant manifold  $\{a \in A_* : \omega(a) \cdot v = 0 \ \forall v \in L\}$ . The neighborhood  $A_L$  is chosen in such a way that all its points have frequency vectors  $\omega(a)$  which do not have any other resonance of order  $\leq N(\varepsilon)$  except those with the vectors of  $L$ . Precisely, if  $a \in A_L$ , then  $a$  is at distance  $\leq \varepsilon^{c_1}$  from the  $L$ -resonant manifold and  $|\omega(a) \cdot v| > \varepsilon^{c_2}$  for all  $v \in \mathbb{Z}^n \setminus L$ ,  $|v| \leq N(\varepsilon)$ , with suitable constants  $c_1, c_2 > 0$  which, for a quasi-convex  $k$ , depend only on the rank of  $L$ . Next, the proof builds, in each set  $c^{-1}(A_L)$ , a semiglobal normal form adapted to  $L$ , of the form

$$k + \varepsilon f_L + \varepsilon e^{-N(\varepsilon)} r_L. \quad (15)$$

Here,  $f_L$  and  $r_L$  are functions which depend on  $\varepsilon$  and are uniformly bounded in  $\varepsilon$ , and  $f_L$  “has all its

Fourier components zero except those in the lattice  $L$ ." Even though the last statement can be given a semiglobal meaning (see the Appendix), the simplest way of expressing it is that the local representative  $f_L^{\text{loc}}$  of  $f_L$  in each system of generalized action-angle coordinates has Fourier spectrum contained in  $L$ , namely,

$$f_L^{\text{loc}}(a, \alpha, p, q) = \sum_{v \in L} (f_L^{\text{loc}})_v(a, p, q) e^{iv \cdot \alpha}. \quad (16)$$

We call function (15) the  $L$ -resonant Nekhoroshev normal form if  $L \neq \{0\}$  and the nonresonant Nekhoroshev normal form if  $L = \{0\}$ . We stress that these normal forms are "semiglobally" defined in the manifold  $c^{-1}(A_L)$  and can be used to analyze the dynamics even if motions are not confined to a chart domain.

In order to understand the dynamics on times up to the Nekhoroshev time  $T_{\text{exp}}$ , one may ignore the exponentially small remainder in (15) and consider the truncated Hamiltonian

$$k + \varepsilon f_{\{0\}} \quad (17)$$

which is defined in  $c^{-1}(A_L)$  and produces an approximate description of that dynamics. There are a few cases to consider.

1. *Motions in nonresonant flowers.* In a nonresonant Nekhoroshev normal form, the local representatives of the function  $f_{\{0\}}$  are independent of the angles  $\alpha$ . Therefore, the truncated Nekhoroshev nonresonant normal form  $k + \varepsilon f_{\{0\}}$  can be viewed as a function defined in the open subset  $\tilde{c}^{-1}(A_L)$  of the base  $B$  of the fibration by the invariant tori, which is a Poisson manifold.

In  $B$ , the Hamiltonian vector field  $X_{k+\varepsilon f_{\{0\}}}^B$  of the truncated nonresonant normal form is tangent to the symplectic leaves  $a = \text{const}$  and, since  $k$  is a Casimir of  $B$ , equals  $X_{\varepsilon f_{\{0\}}}^B$  and thus  $\varepsilon X_{f_{\{0\}}}^B$ . Thus, up to a time-reparametrization by  $\varepsilon$ , the truncated nonresonant normal form defines a family of Hamiltonian systems, one on each symplectic leaf with Hamiltonian the restriction of the function  $f_{\{0\}}$  to that symplectic leaf, parametrized by the

actions  $a$ . (Note: In reality, the actions do not stay constant and motions do not remain in a flower and in a symplectic leaf. However, since in our hypotheses the symplectic leaves are the fibers of a fibration, they may all be identified to a given one). In general, these restricted Hamiltonian systems will not be integrable, and their flow will be chaotic. In special cases they – or some of them, on certain symplectic leaves – are integrable and the motion along the symplectic leaves is regular. In other special cases, these restricted Hamiltonian systems are, in certain regions of certain symplectic leaves, small perturbations of "intermediate" integrable systems and, again, motions are regular (possibly on shorter time-scales, determined by the properties of the intermediate integrable system).

A special case in which the motion along the symplectic leaves of the nonresonant flowers is always regular is, under some genericity conditions on the perturbation, when the symplectic leaves are two-dimensional ( $d = n + 1$ , any  $n \geq 1$ ). Indeed, in such a case the Hamiltonian systems given by the restrictions of  $f_{\{0\}}$  to the symplectic leaves have one degree of freedom and are integrable. More precisely, they describe motions that, on each symplectic leaf, follow the level sets of the function  $f_{\{0\}}|_a = \text{const}$ . This function differs by quantities small with  $\varepsilon$  from the average  $f_0$  of the perturbation  $f$  on the invariant tori (namely, the function whose local representatives are the averages over the angles of the local representatives of  $f$ ; see also the Appendix). Thus, the flow of the truncated nonresonant normal form on each symplectic leaf follows approximately the level curves of the restriction of  $f_0$  to it, which can often be computed explicitly. If, for instance, the restriction of  $f_0$  to the symplectic leaves has only isolated critical points, then almost all these level sets are regular curves. In particular, if the symplectic leaves are compact, most of them are closed curves and motions described by the truncated normal forms on the symplectic leaves are close to periodic motions.

In the literature, this fact is often expressed by considering the action-angle coordinates of the integrable systems given by  $f_{\{0\}}|_a = \text{const}$

and saying that, in these cases, Nekhoroshev theorem produces the stability of all actions.

2. *Motions in resonant flowers.* In a resonant normal form, the local representatives of the function  $F_L$  depend also on  $r = \text{rank}(L)$  linear combinations of angles (the so called “slow angles”). As a consequence, even in the approximation provided by the truncated normal form, the motion on the symplectic leaves is coupled to that in the petals. (Note: The truncated normal form (17) can be viewed as defined in a sub-bifibration which has action space and petals of dimension  $r$  and has the same symplectic leaves as  $B$ ). Even in the simple case with two-dimensional symplectic leaves, such a system has more than one degree of freedom and may be not integrable. Moreover, the projection of motions on the symplectic leaves need not stay close to the level curves of a regular function on the symplectic leaves and could erratically – even though slowly – wander through extended, two-dimensional regions of the symplectic leaves.

The presence of such slow chaotic motions along the symplectic leaves is demonstrated numerically and is theoretically well understood when  $d = 3$  and  $n = 2$ . In that case, the only resonance of interest are single resonances, with resonant lattice of rank 1. The truncated normal form depends only on one “slow angle” (a certain linear combination of the two angles) and leaves a linear combination of the two actions constant. Hence, it can be viewed as describing a perturbation of a superintegrable system with one-dimensional action space and petals and two-dimensional symplectic leaves, hence, with two degrees of freedom. In typical situations, such a system is a pendulum-like system on the symplectic leaves which is periodically forced by the motion of the “slow” angle. Hence, the separatrices of the pendulum move (approximately) periodically in time, with speed of order  $\varepsilon$ . In the region of the symplectic leaf away from the separatrices, there is an approximate integral of motion (an adiabatic invariant) and motions take place along its level curves and are regular. However, if a

motion enters the region swept by a separatrix, when it crosses the separatrix the adiabatic invariant undergoes a small jump that makes the system move to a different level curve. The accumulation of these jumps produces a slow diffusive-like behavior that takes motions to fill an extended region of the symplectic leaf; an important aspect is that these chaotic motions are present no matter how small  $\varepsilon$  is and that the amplitude of the visited region does not go to zero for  $\varepsilon \rightarrow 0$ . We will refer to this behavior as “slow extended chaos.” These phenomena are explained and described by a theory based on the adiabatic theory developed by Neishtadt (Neishtadt 1987) and others in the 1980s; see (Arnold et al. 2006; Guzzo 2007; Neishtadt and Sidorenko 2004).

Very little is presently known in cases with invariant tori and/or symplectic leaves of higher dimensions. However, reference (Guzzo 1999) remarks that, if in a region of the phase space the average  $h_0 = k + \varepsilon f_0$  of the hamiltonian is (a small perturbation of) an integrable and convex system, then within the timescale  $T_{\text{exp}}$ , slow large-amplitude chaotic phenomena are to be expected only in low-order resonances. Specifically, for any given positive  $\varepsilon_1 < \varepsilon_*$ , for all  $\varepsilon < \varepsilon_1$ , motions are regular in all those resonant flowers whose resonant lattices do not have generators of norm  $< \ln \frac{\varepsilon_*}{\varepsilon_1}$ . The reason for this is the exponential decay of the Fourier components of real analytic functions. This situation is exceptional but is verified, for instance, in the spatial circular restricted 3-body problem with small mass-ratio and in the system describing the motion of an asteroid in the main belt under the perturbation of the planets. For further details and references, see (Guzzo 2007).

### KAM Theory

As already mentioned, the techniques of KAM theory allow to prove the persistence of strongly nonresonant (e.g., with Diophantine frequencies) *lagrangian*, not of *isotropic*, invariant tori of a completely integrable system into lagrangian invariant tori of the perturbed one. (Note: An exception are the isotropic tori with a hyperbolic normal behavior, which is not the case of those of

a superintegrable system). Moreover, this construction works only when the frequencies of the lagrangian tori of the unperturbed system change “as much as possible” from torus to torus (as ensured, e.g., by Kolmogorov’s non-degeneracy condition). Accordingly, the application of KAM theory to superintegrable systems rests on the possibility of building, using the perturbation, an “intermediate” nondegenerate completely integrable system to which KAM theory can then be applied. This is of course possible only for special perturbations and, often, only in certain regions of the phase space. This procedure, which is called “removal of the degeneracy” of the superintegrable system by means of the perturbation, was envisaged and developed by Arnold in order to apply the newly created KAM theory to perturbations of planetary systems (Arnold 1963). (Note: The idea of removing the degeneracy by means of the perturbation comes from the classical, still formal, pre-KAM perturbation theory; see Sect. 18 of Born (1960).)

Assume that

- (i) The superintegrable Hamiltonian  $k$  satisfies Kolmogorov nondegeneracy condition.
- (ii) In a certain subset  $c^{-1}(U)$ ,  $U \subseteq A$ , of the phase space, there exists an  $\varepsilon$ -dependent symplectic diffeomorphism  $C_\varepsilon : c^{-1}(U) \rightarrow C_\varepsilon(c^{-1}(U)) \subseteq M$  which is close to the identity for small  $\varepsilon$  and is such that

$$(k + \varepsilon f) \circ C_\varepsilon = k + \varepsilon \bar{G} + \varepsilon^s \mathcal{R} \quad (18)$$

with some  $s > 1$  and  $\varepsilon$ -dependent functions  $\bar{G}$  and  $\mathcal{R}$  (uniformly bounded as  $\varepsilon \rightarrow 0$ ) such that:

- $\bar{G}$  is constant on the (images under  $C_\varepsilon$ ) of the flowers.
- For each fixed  $\varepsilon$ , the restriction of  $\bar{G}$  to (the image under  $C_\varepsilon$  of) each symplectic leaf defines a completely integrable system (with  $d - n$  frequencies) which satisfies Kolmogorov nondegeneracy condition.

Under these conditions, the truncated Hamiltonian  $k + \varepsilon \bar{G}$  (sometimes called “intermediate”) is a completely integrable Hamiltonian with

$d$  degrees of freedom that satisfies Kolmogorov’s nondegeneracy condition. Applying KAM theorem to (18), regarded as an  $\varepsilon^s$ -perturbation of the intermediate system, gives the existence of a set of large measure of invariant  $d$ -dimensional tori. Each of these KAM tori is close to the Cartesian product of a strongly nonresonant  $n$ -dimensional invariant torus of the superintegrable system and of a strongly nonresonant  $(d - n)$ -dimensional torus of the restriction to a symplectic leaf of the system of hamiltonian  $\bar{G}$ . Motions on these tori take place with speed of order one on the  $n$ -dimensional subtorus and with speed of order  $\varepsilon$  on the  $(d - n)$ -dimensional one.

This KAM result can be seen as a further description of motions on certain nonresonant flowers of the Nekhoroshev theory. On each strongly nonresonant flower (which form a Cantor-like subset of large measure of the (approximately) nonresonant flowers considered within the realm of Nekhoroshev theorem), there is a Cantor-like set of large measure of invariant tori on which motions are, for all times, linear. In conclusion, under the stated hypotheses, *there is a Cantor set of large measure of coisotropic flowers on each of which the perturbation creates a Cantor set of large measure of invariant  $d$ -dimensional tori.*

**Remark** In practice, the intermediate nearly integrable Hamiltonian (18) is often built via averaging techniques, so that  $\bar{G} = f_0 + O(\varepsilon^2)$ , with  $f_0$  the average of  $f$  (see the Appendix).

### Some Applications

We describe now a few applications of the above results and techniques.

1. *Perturbations of the Euler top* fall in the best understood case  $d = 3$ ,  $n = 2$ . Recall from section *Examples* that the two actions  $a = (a_1, a_2 = \|m^s\|)$  fix the kinetic energy  $k$  and the norm of the angular momentum vector and that the symplectic leaves are two-dimensional spheres whose points represent the direction of  $m^s$  in space, restrict to motions with nonzero angular momentum, and define  $\mu := \frac{m^s}{\|m^s\|}$ . Then, the perturbed Hamiltonian has the form  $k(a) + \varepsilon V(a, \alpha, \mu)$ , where  $V$  (that we assume to be real analytic) is the potential of the external forces

acting on the body. The kinetic energy is a quasi-convex ((Benettin and Fassò 1996), section 7) and nondegenerate (Mazzocco 1997) function of the two actions.

This system has been studied via Nekhoroshev theorem in Benettin and Fassò (1996), which focused on the case of a dynamically symmetric body but explains how the results apply to the generic case. (Note: In that work, the emphasis is on “fast” motions of the body and the small parameter is not the size of the perturbation but the inverse of the norm  $w$  of the body’s angular velocity. The correct relationship between the two parameters is  $w \sim \varepsilon^{-1/2}$  (Benettin et al. 2004) and not, as erroneously hinted in Benettin and Fassò (1996),  $w \sim \varepsilon^{-1}$ ). Exclude a neighborhood of the boundary of the action space (which contains the projections of the following motions of the Euler top: equilibria, “stationary rotations” with the angular momentum aligned with one of the principal axes of inertia, and motions in the stable and unstable manifolds of the stationary rotations about the middle inertia axis). The symplectic leaves are compact, and Proposition 7 gives stability of the two actions  $a_1$  and  $a_2$  for times  $T_{\text{exp}}$ , with  $\beta = \frac{1}{4}$ . The properties of the motion along the symplectic leaves on the timescale  $T_{\text{exp}}$  depend on the resonance properties of the unperturbed motion. Out of resonance,  $\mu$  moves slowly, with speed order  $\varepsilon$ , along the level curves of the average of  $V$ , which are generically closed curves. The motion resembles a motion of the Euler top, but with the direction of the angular momentum vector moving slowly and regularly in space. In resonances of a not-too-high order, instead,  $\mu$  may wander chaotically through extended regions of the unit sphere, with positive Lyapunov exponents. Numerical illustrations of these behaviors are given in Benettin et al. (2002).

Perturbations of the Euler top have been studied via KAM theorem in Hanßmann (1997); Mazzocco (1997), who proved the existence of a Cantor set of large measure of strongly nonresonant flowers in which there is a Cantor set of three-dimensional invariant

KAM tori. These KAM tori are close to the Cartesian product of a two-dimensional invariant torus of the Euler top and of a closed level curve of the average of the restriction of  $V$  to a symplectic leaf.

2. *The perturbed Euler top near stationary rotations.* Even though we have not stressed this fact so far, Nekhoroshev theory applies also to the study of perturbations of integrable and superintegrable systems with  $n$  frequencies in neighborhoods of certain singularities of the foliations by their invariant tori, particularly normally elliptic invariant tori of dimension  $< n$  (an example is elliptic equilibria (Fassò et al. 1998; Niederman 1998)). Singularities of this type, in the Euler top, are the stationary rotations around the minimum and maximum inertia axes, which belong to singular flowers located at the boundary of the action space and which have regular symplectic leaves, diffeomorphic to two-dimensional spheres, but one-dimensional petals. Motions of perturbations of the Euler top with the angular momentum vector, initially, nearly aligned with one such inertia axis (“gyroscopic rotations”) have been studied via the Nekhoroshev theory in Benettin et al. (2004) (here too, in the dynamically symmetric case, but the results extend to the generic case). For the entire timescale  $T_{\text{exp}}$  (here too with  $\beta = 1/4$ ), the angular momentum vector remains, in the body, near the inertia axis and the general patterns of motions are the same as for the nongyroscopic motions described above, with however an important difference: In resonance, the direction of the angular momentum vector in space moves still chaotically (with positive Lyapunov exponents) but, instead of wandering through extended regions of the unit sphere, remains close to the level curves of a function. This makes precise the idea that gyroscopic rotations are more stable than generic motions. Numerical investigations of these phenomena are reported in Benettin et al. (2006). The case of a flat body (which corresponds to the resonance 1:1) presents some peculiarities studied and explained, through a suitable version of the adiabatic theory, in Benettin et al. (2008b).

3. *Perturbations of central forces.* For the point in a central force field, in a region in which motions are all bounded,  $d = 3$ ,  $n = 2$  and the symplectic leaves are two-dimensional spheres. Reference (Bambusi et al. 2018) proves that, for all real analytic central potentials (with the exception of the Keplerian and of the harmonic ones, which by Bertrand theorem are well known to be the only ones for which all bounded motions are periodic, and hence maximally superintegrable), the Hamiltonian is quasiconvex and applies Nekhoroshev theorem for superintegrable systems to prove stability of the two actions over the timescale  $T_{\text{exp}}$  with  $\beta = 1/4$ . The motion along the symplectic leaves describes the variation of the spatial orientation of the plane of the orbit, which might undergo slow extended chaotic movements in resonance, but has not been investigated so far.
4. *Stability of planetary systems.* As already noticed, in Nekhoroshev (1977) formulated his own theorem in a form that applies (semilocally) to superintegrable systems and applied it to study the stability of a “planetary system,” namely, the  $(n + 1)$ -body system described in the Introduction of this section ( $n$  planets of small mass and a star of mass 1, in a star-centric frame). Such system is regarded as a small perturbation of the superintegrable system formed by the  $n$  non-interacting Keplerian systems star-planet in a subset of its phase space that corresponds to planetary motions: a subset where each planet moves on a Keplerian ellipse with sufficiently small eccentricities and inclinations, all ellipses have sufficiently different semimajor axes and are run in the same direction (“domain of planetary motions”). The  $n$  actions are the semimajor axes of these Keplerian ellipses, and the  $n$  angles are mean anomalies along them. The points in the symplectic leaves fix inclinations, orientations, and eccentricities of these ellipses. The Hamiltonian of this superintegrable system is convex.

Each such domain of planetary motions is an open subset of the phase space that intersects the symplectic leaves in open subsets which can be covered by a single system of

generalized action-angle coordinates, and Nekhoroshev’s original semilocal formulation of his theorem can be used. Nevertheless, there is the need to control that motions remain in the domain of planetary motions (in particular, it is necessary to exclude collisions). To this end, Nekhoroshev employs a well-known argument by d’Alembert, which rests on the conservation of the total angular momentum of the system, coupled to the conservation of the actions provided by his theorem. In short, the norm of the angular momentum vector of a Keplerian system is, for given semimajor axis, maximal on circular orbits. Since initially the mutual inclinations of the orbits are small and the orbits are run in the same direction, initially the norm of the total angular momentum vector is close to its maximum value for the given initial energy. Nekhoroshev theorem gives the approximate conservation of the semimajor axis and thus of the energy of each planet (which in a Keplerian system depends only on the semimajor axis). This, together with the fact that the norm of the total angular momentum is constant, implies that also the norm of the angular momentum of each planet remains nearly constant and close to its maximum. In conclusion, for the timescale  $T_{\text{exp}}$  (with  $\beta = \frac{1}{2n}$ ), the planets keep moving on nearly circular and nearly coplanar orbits (run of course in the same direction). This study has been repeated in Niederman (1996).

The application of KAM theory to planetary systems, envisaged and began by Arnold (1989), predates Nekhoroshev’s one and is still a subject of investigation. Restrict to a domain of planetary motions. In a system of generalized action-angle coordinates, the Hamiltonian has the form  $k(a) + \varepsilon V(a, \alpha, p, q)$  where the actions  $a$ ’s are the semimajor axes of the Keplerian ellipses, the  $\alpha$ ’s are the mean anomalies of these Keplerian motions, and the  $(p, q)$ ’s fix inclinations, orientation, and eccentricities of these ellipses. The unperturbed part is Kolmogorov nondegenerate (as function of the  $n$  actions  $a$ ), but building a completely integrable “intermediate” Hamiltonian faces the difficulty that the restriction to the symplectic



leaves of the average  $V_0$  of  $V$  is not completely integrable. However, the system with Hamiltonian  $V_0|_a = \text{const}$  has an elliptic equilibrium which corresponds to all planets moving on coplanar circular orbits. In a small neighborhood of that equilibrium, say of size  $\eta$ , the system is a small perturbation of its linearization. In the absence of resonances of order  $\leq 4$  among the frequencies of this elliptic equilibrium,  $V_0$  can be expanded in a Birkhoff normal form  $V_{00} + \eta V_{01} + O(\eta^2)$  where  $V_{00}$  is the Hamiltonian of the linearization of  $V_0$  at the equilibrium and describes a completely integrable system with fixed frequencies, hence degenerate (it is a quadratic form in Birkhoff Cartesian-like coordinates). Arnold suggested using the higher order terms of the Birkhoff normal forms to remove the degeneracy of the quadratic part. The “fundamental theorem” in Arnold (Arnold 1963) is a formulation of KAM theorem for perturbations of a system with Hamiltonian  $k(a) + \varepsilon V_{00} + \varepsilon \mu V_{01} + \dots$  of this type, which depends on two small parameters. An improved version of this KAM result has been given in Chierchia and Pinzari (2010). In order to prove the existence of a large set of invariant tori of dimension  $3n$  in the (heliocentric)  $n + 1$  planetary system, it remained to prove the nondegeneracy. Arnold did it for  $n = 2$  planets and under the additional restriction that the system moves on a plane (hence the tori have dimension  $2n = 4$ ). He claimed that the same is true in general, but the proof of this fact required many efforts and was achieved by Robutel (Robutel 1995) for the spatial case with 2 planets and then by Féjóz (Féjóz 2004) and Chierchia and Pinzari (Chierchia and Pinzari 2011) for the general case (the latter paper contains further details on the problem, which is very subtle). This allows to conclude that, in a domain of planetary motions, there is a set of a large measure of KAM tori of dimension  $3n$ , which are close to the  $n$ -dimensional tori of the system formed by the  $n$  uncoupled star-planet systems.

5. *Other examples* of application of KAM theory to perturbations of superintegrable systems can be found in Hanßmann (2015).

## Conclusion

The study of perturbations of superintegrable systems has been a motivation for the development of KAM and Nekhoroshev theories since their beginning.

KAM theory for superintegrable systems relies on the use of the perturbation to remove the degeneracy by identifying, in some region of the phase space, a suitable “intermediate” non-degenerate integrable system whose Lagrangian tori are then proved to survive the (remaining part of the) perturbation by means of KAM theorem. Usually, this can be done working semilocally, in a system of generalized action-angle coordinates of the superintegrable system. The consideration of the semiglobal superintegrable structure may add comprehension.

Nekhoroshev theory for superintegrable systems requires instead in an essential way a comprehension of the geometric structure of these systems and the construction of semiglobal normal forms. The statement given in Proposition 7 can be easily generalized assuming steepness, instead of convexity or quasi-convexity, of the superintegrable Hamiltonian.

The most interesting feature in the dynamics of perturbations of superintegrable systems is, probably, the richness of behaviors of the projection of motions along the symplectic leaves on the exponentially long Nekhoroshev timescale, which is presently understood only in special cases. When the symplectic leaves are two-dimensional, the regularity or chaoticity of these motions depends on the (non)resonance properties of the unperturbed motion. The development of chaotic motions, and their “slow but extended” pattern, is understood only when, in addition,  $n = 2$ . Further studies are advisable for the case of two-dimensional symplectic leaves and  $n > 2$  and for the case of symplectic leaves of dimension  $> 2$ .

## Appendix: Semiglobal Normal Forms

1. *Fourier series and averages on a bifibration.* In this Appendix, we describe the bases on which rests the construction of semiglobal normal forms.

The setting is that of a bifibration  $M \xrightarrow{y} B \xrightarrow{\tilde{c}} A$  with  $\dim M = d \geq 2$ ,  $\dim A = n$  with some  $1 \leq n \leq d - 1$ , and no monodromy.

First, we make a remark. It was proven in Cushman (1984); Moser (1970) that the construction of normal forms for vector fields on a manifold whose flow is *periodic* is a geometric operation that does not require the consideration of any coordinate system. This is because, for a periodic flow, spatial averaging over the orbits can be replaced with time averaging, which uses only the flow and produces objects globally defined in the manifold (or at least, in the subset of it in which the flow is periodic, with exclusion of the equilibria, where the normal form is of interest). The situation is different for vector fields on manifolds which, like superintegrable systems with  $n > 1$  frequencies, have dynamics linear on tori of dimension  $> 1$ , because averaging over them cannot be replaced by time averaging. (Note: However, the construction of normal forms for “maximally resonant” tori, which carry periodic motions, can be done via time averaging. Blaom’s proof (Blaom 2001) of the Nekhoroshev theorem for superintegrable system that we mentioned above uses this technique and need not exploit the symplectic geometry of the bifibration). The *semiglobal* existence, regularity, and computation of spatial averages and normal forms in a bifibration without monodromy has been proven in Fassò (1995) and rests on the fact that sets of flowers without monodromy are  $\mathbb{T}^n$ -principal bundles and there is a (semiglobal) Fourier series defined on them.

**Proposition 8** *Consider a bifibration  $M \xrightarrow{y} B \xrightarrow{\tilde{c}} A$  with connected and simply connected action space  $A$ , and let  $n := \dim A \geq 1$ . Choose an atlas with generalized action-angle coordinates for  $M$  which has all matrices  $Z$  equal to*

*the identity. Then, for any function  $f : M \rightarrow \mathbb{R}$ , there exist functions  $f_v : M \rightarrow \mathbb{R}$ ,  $v \in \mathbb{Z}^n$ , which are such that*

- (i)  $f = \sum_{v \in \mathbb{Z}^n} f_v$ .
- (ii) *In any chart  $(a, \alpha, p, q)$  of the considered atlas,  $\{f_v, a_j\} = iv_j f_v$  for all  $j = 1, \dots, n$ .*

*The functions  $f_v$  are independent of the considered atlas; it is only their labeling with the integer vectors that depends on the choice of the atlas.*

**Proof** The representative  $f^{\text{loc}}$  of  $f$  in a system of generalized action-angle coordinates  $\mathcal{T} = (a, \alpha, p, q)$ . can be expanded in Fourier series  $f^{\text{loc}} = \sum_{v \in \mathbb{Z}^n} f_v^{\text{loc}}$ , with  $f_v^{\text{loc}}(a, p, q, \alpha) = \hat{f}_v^{\text{loc}}(a, p, q) e^{iv \cdot \alpha}$ ,  $\hat{f}_v^{\text{loc}} = (2\pi)^{-n} \int_{\mathbb{T}^n} f^{\text{loc}} e^{-iv \cdot \alpha} d\alpha$  being the “Fourier components.” The same can be done for the representative  $f^{\text{loc}'}$  of  $f$  in any other system of generalized action-angle coordinates  $\mathcal{T}' = (a', \alpha', p', q')$  whose domain intersects that of  $\mathcal{T}$ . By (7), the matching condition  $f \circ \mathcal{T} = f^{\text{loc}'} \circ \mathcal{T}'$  gives, in the intersection of the domains,

$$\begin{aligned} f_{Z^T v}^{\text{loc}}(a, p, q) &= f_v^{\text{loc}'} \left( Z^{-T} a + z, \mathcal{P}(a, p, q), \right. \\ &\quad \left. Q(a, p, q) \right) e^{iv \cdot \mathcal{F}(a, p, q)} \quad \forall v \in \mathbb{Z}^n. \end{aligned} \quad (19)$$

It follows that, in an atlas  $\mathcal{A}$  with all transition matrices  $Z = \mathbb{I}$ , for each  $v \in \mathbb{Z}^n$  the  $v$ -harmonics of the local representatives of  $f$  match (namely,  $f_v \circ \mathcal{T} = f_v^{\text{loc}'} \circ \mathcal{T}'$  in the intersection of the domains) and hence are the local representatives of a function  $f_v : M \rightarrow \mathbb{R}$ . (This does not however happen for the Fourier components.) Statement i. is now obvious. Statement ii. follows from the fact that  $\{f_v, a_j\} = \frac{\partial f_v^{\text{loc}}}{\partial \alpha_j} \circ \mathcal{T} = iv_j f_v$ . Consider now another atlas  $\mathcal{A}'$  with all transition matrices  $Z = \mathbb{I}$ . The transition functions between the charts of the two atlases have the form (7) with a given matrix  $Z$ , the same for all pairs of charts. Therefore, since  $\mathcal{A}$  is connected, it follows from (19) that, for each  $v$ , the harmonic of  $f$  labeled with  $v$  in the atlas  $\mathcal{A}'$

coincides with the harmonic of  $f$  labeled with  $Z^T v$  in the atlas  $\mathcal{A}$ .

Thanks to this fact, it is possible to space-average functions on a bifibration. In the hypotheses and with the notation of Proposition 8, given any subset  $L$  of  $\mathbb{Z}^n$ , define the  $L$ -average of a function  $f: M \rightarrow \mathbb{R}$  as

$$\Pi_L f := \sum_{v \in L} f_v : M \rightarrow \mathbb{R}.$$

Here too, only the labeling by  $L$  depends on the choice of the atlas. If  $L$  is a lattice of  $\mathbb{R}^n$  of positive rank  $r$ ,  $0 \leq r \leq n$ , then  $\Pi_L f$  is a function which, on each invariant torus, coincides with the spatial average of  $f$  over the  $(n - r)$ -dimensional subtori orthogonal to  $L$ .

In particular, the restriction of  $\Pi_{\{0\}} f = f_0$  to each invariant torus is the spatial average of  $f$  over that  $n$ -dimensional torus. Since its local representatives are independent of the angles,  $f_0$  is the lift of a function  $\tilde{f}_0$  on the base  $B$ ,  $f_0 = \tilde{f}_0 \circ \gamma$ . The Hamiltonian vector field of  $f_0$  in  $M$  is tangent to the flowers (its local representatives are independent of the angles  $\alpha$ ) and that of  $\tilde{f}_0$  in  $B$  is tangent to the symplectic leaves.

**2. Symplectic diffeomorphisms.** An effective way of constructing the symplectic diffeomorphisms which lead to the normal forms on a bifibration is by realizing them via the flow of Hamiltonian vector fields. This is the so-called Lie method, which even though usually implemented in coordinates is in fact fully geometric and then suitable to the purpose.

Let  $M$  be a symplectic manifold and  $\chi: M \rightarrow \mathbb{R}$  be a smooth function. Denote by  $\Phi^\chi$  the flow of the Hamiltonian vector field  $X_\chi$ . The time-one map  $\Phi_1^\chi$  of this flow is symplectic, and if  $\chi$  is small enough (in some norm), then there exists an open nonempty subset  $M_1$  of  $M$  such that  $\Phi_1^\chi: M_1 \rightarrow \Phi_1^\chi(M_1)$  is a diffeomorphism. Moreover, if  $\chi$  is small, then  $\Phi_1^\chi$  is near the identity, in the sense that  $\|\Phi_1^\chi - \text{id}\| = O(|\chi|)$  (in suitable norms).

The “Lie transform” generated by  $\chi$  is the transformation of functions  $f \mapsto f \circ \Phi_1^\chi$ . Since  $\frac{d}{dt}(f \circ \Phi_t^\chi) = -L_{X_\chi} f = \{f, \chi\} \circ \Phi_t^\chi$ ,

$$f \circ \Phi_1^\chi = f + \mathcal{R}_1^\chi(f) = f + \{f, \chi\} + \mathcal{R}_2^\chi(f) \quad (20)$$

with  $\mathcal{R}_1^\chi(f) = \int_0^1 \{f, \chi\} \circ \Phi_t^\chi dt$  and  $\mathcal{R}_2^\chi(f) = \int_0^1 \int_0^t \{\{f, \chi\}, \chi\} \circ \Phi_s^\chi \circ \Phi_t^\chi ds dt$ . These expressions provide approximations of  $f \circ \Phi_1^\chi$  up to remainders  $\mathcal{R}_s^\chi$  which are, in an obvious sense, of order  $|\chi|^s$ ,  $s = 1, 2$ .

Moreover, if  $\chi^{\text{loc}}$  and  $f^{\text{loc}}$  are the local representatives of  $\chi$  and  $f$  in a symplectic chart of  $M$ , then the local representatives of the Hamiltonian vector field  $X_\chi$ , of the diffeomorphism  $\Phi_1^\chi$  and of the maps  $\mathcal{R}_s^\chi(f)$  ( $s = 1, 2$ ) are, respectively,  $X_{\chi^{\text{loc}}}$ ,  $\Phi_1^{\chi^{\text{loc}}}$  and  $\mathcal{R}_s^{\chi^{\text{loc}}}(f^{\text{loc}})$ . Thus, estimates, etc., can be performed using local representatives.

**3. The semiglobal homological equation.** Consider now a perturbation  $k + \varepsilon f$  of a superintegrable system with Hamiltonian  $k$  on a bifibration  $M \xrightarrow{\gamma} B \xrightarrow{\tilde{\gamma}} A$  with trivial monodromy. Choose an atlas with generalized action-angle coordinates having all matrices  $Z = \mathbb{R}$ , and as explained in the subsection on the semiglobal normal form of the section on Nekhoroshev theorem, let  $\omega = \frac{\partial k}{\partial a}: A \rightarrow \mathbb{R}^n$  be the frequency map of  $k$ .

The basic step in constructing a normal form for  $h + \varepsilon f$  “adapted” to a “resonant” subset  $L$  of  $\mathbb{R}^n$  consists in the elimination, at some order in  $\varepsilon$ , of all the Fourier components of the perturbation  $f$  except those belonging to  $L$ . In typical cases, including the proof of Nekhoroshev theorem,  $L$  consists of a “resonant” sublattice  $L^*$  of  $\mathbb{Z}^n$  and also, if a cut-off  $N$  is used, of all the integer vectors of norm  $\text{norm} > N: L = L^* \cup \{v \in \mathbb{Z}^n: |v| \geq N\}$ . Therefore, the complement  $\mathbb{Z}^n \setminus L$  is a finite set; we make here this assumption. The normal form can be built in the subset  $c^{-1}(R_L)$  of  $M$ , with

$$R_L = \{a \in A: \omega(a) \cdot v \neq 0 \quad \forall v \in \mathbb{Z}^n \setminus L\}.$$

(However, in order to prevent the normal form to diverge at the points over the boundary of  $R_L$  and to produce estimates on it, it is necessary to restrict somehow  $R_L$ .)

If  $\chi$  is any function, then by (4.2)  $(k + \varepsilon f) \circ \Phi_1^\chi = k + \{k, \chi\} + \varepsilon f + \mathcal{R}_1^\chi(\varepsilon f) + \mathcal{R}_2^\chi(k)$ . Thus, if  $\chi$  satisfies the “homological equation”

$$\{k, \chi\} = \varepsilon (\Pi_L f - f), \quad (21)$$

then

$$(k + \varepsilon f) \circ \Phi_1^\chi = k + \varepsilon \Pi_L f + R_1^\chi(\varepsilon f) + R_2^\chi(k)$$

is, up to the remainders, a normal form for  $k + \varepsilon f$  adapted to the resonant set  $L$  and is semiglobally defined in  $c^{-1}(R_L)$ .

The “semiglobal” homological Eq. (21) (which is written in an intrinsic way, rather than in local coordinates) can be solved, exactly as in the semi-local perturbation theory that uses generalized action-angle coordinates, by Fourier series techniques. Indeed,  $\{k, \chi\} = \sum_{j=1}^n \frac{\partial k}{\partial a_j} \{a_j, k\} = \sum_{j=1}^n \sum_{v \in \mathbb{Z}^n} \omega_j \{a_j, \chi_v\}$  and, since by item ii. of Proposition 8  $\{a_j, \chi_v\} = -iv_j \chi_v$ , Eq. (21) can be written as

$$\sum_{v \in \mathbb{Z}^n} i\omega \cdot v \chi_v = \sum_{v \notin L} f_v$$

and, in  $c^{-1}(R_L)$ , has the solution

$$\chi = \sum_{v \notin L} \frac{f_v}{i\omega \cdot v}$$

(which is well defined, given that  $\mathbb{Z}^n \setminus L$  is finite).

This shows that normal forms can be built semiglobally.

**Acknowledgments** The author wishes to thank Giuseppe Gaeta for his kind invitation to write this item, the Springer Reference Team for their helpful collaboration, and his colleagues Massimiliano Guzzo and Gabriella Pinzari for some very useful conversations. The author has been partially supported by the MIUR-PRIN project 20178CJA2B *New Frontiers of Celestial Mechanics: theory and applications*.

## Bibliography

- Arnold VI (1963) Small denominators and problems of stability of motion in classical and celestial mechanics. *Russ Math Surv* 18:85–191
- Arnold VI (1989) *Mathematical methods of classical mechanics*, 2nd edn. Springer, New York

- Arnold VI, Givental AB (2001) *Symplectic geometry*. In: *Dynamical systems*, Encyclopaedia Math Sci 4, vol IV. Springer, Berlin, pp 1–138
- Arnold VI, Kozlov VV, Neishtadt AI (2006) *Mathematical aspects of classical and celestial mechanics*. [Dynamical systems III], Encyclopaedia of mathematical sciences, 3, 3rd edn. Springer-Verlag, Berlin
- Bambusi D, Fusè A, Sansottera M (2018) Exponential stability in the perturbed central force problem. *Regul Chaotic Dyn* 23:821–841
- Bates LM (1988) Examples for obstructions to action-angle coordinates. *Proc Roy Soc Edinburgh* 110A:27–30
- Bates L, Fassò F (2016) No monodromy in the champagne bottle, or singularities of a superintegrable system. *J Geom Mech* 8:375–380
- Benettin G (2005) The elements of Hamiltonian perturbation theory. In: Benest D, Froeschlé C, Lega E (eds) *Hamiltonian systems and fourier analysis*. Cambridge Scientific Publishers, Cambridge, pp 1–98
- Benettin G, Fassò F (1996) Fast rotations of the rigid body: a study by Hamiltonian perturbation theory. Part I. *Nonlinearity* 9:137–186
- Benettin G, Cherubini AM, Fassò F (2002) Regular and chaotic motions of the fast rotating rigid body: a numerical study. *Discrete Contin Dyn Syst B* 2:521–540
- Benettin G, Fassò F, Guzzo M (2004) Long-term stability of proper rotations of the Euler perturbed rigid body. *Commun Math Phys* 250:133–160
- Benettin G, Fassò F, Guzzo M (2006) Long term stability of proper rotations and local chaotic motions in the perturbed Euler rigid body. *Regul Chaotic Dyn* 11:1–17
- Benettin G, Guzzo M, Marini V (2008a) Adiabatic chaos in the spin-orbit problem. *Celestial Mech Dyn Astr* 101: 203–224
- Benettin G, Guzzo M, Neishtadt A (2008b) A new problem of adiabatic invariance related to the rigid body dynamics. *Discrete Contin Dyn Syst* 21:959–975
- Blaom AD (2001) A geometric setting for Hamiltonian perturbation theory. *Mem Amer Math Soc* 153:1–112
- Bogoyavlenskij OI (1998) Extended integrability and bi-Hamiltonian systems. *Commun Math Phys* 196: 19–51
- Born M (1960) *The mechanics of the atom*. Frederick Ungar Publishing, New York
- Broer H, Hanßmann H *Hamiltonian perturbation theory (and transition to chaos)*, Encyclopedia of complexity and systems science. Springer
- Chierchia L, Pinzari G (2010) Properly-degenerate KAM theory (following V. I. Arnold). *Discrete Contin Dyn Syst Ser S* 3:545–578
- Chierchia L, Pinzari G (2011) The planetary N-body problem: symplectic foliation, reductions and invariant tori. *Invent Math* 186:1–77
- Chierchia L, Procesi M *Kolmogorov-Arnold-Moser (KAM) theory for finite and infinite dimensional systems*, Encyclopedia of complexity and systems science. Springer

- Cushman RH (1984) Normal form for Hamiltonian vectorfields with periodic flow. In: Sternberg S (ed) *Differential geometric methods in mathematical physics*. Reidel, Dordrecht, pp 125–144
- Cushman R, Bates L (2015) *Global aspects of classical Integrable systems*, 2nd edn. Birkhäuser, Basel
- Dazord P, Delzant T (1987) Le probleme general des variables actions-angles. *J Diff Geom* 26:223–251
- Duistermaat JJ (1980) On global action-angle coordinates. *Commun Pure Appl Math* 33:687–706
- Evans NW (1990) Superintegrability in classical mechanics. *Phys Rev A* 41:5666–5676
- Fassò F (1995) Hamiltonian perturbation theory on a manifold. *Cel Mech Dyn Astr* 62:43–69
- Fassò F (1996) The Euler-Poincot top: a non-commutatively integrable system without global action-angle coordinates. *J Appl Math Phys (ZAMP)* 47:953–976
- Fassò F (1998) Quasi periodicity of motions and complete integrability of Hamiltonian systems. *Ergod Theory Dyn Syst* 18:1349–1362
- Fassò F (2005) Superintegrable Hamiltonian systems: geometry and perturbations. *Acta Appl Math* 87:93–121
- Fassò F, Giacobbe A (2002) Geometric structure of “broadly integrable” Hamiltonian systems. *J Geom Phys* 44:156–170
- Fassò F, Guzzo M, Benettin G (1998) Nekhoroshev-stability of elliptic equilibria of Hamiltonian systems. *Commun Math Phys* 197:347–360
- Fassò F, Giacobbe A, Sansonetto N (2005) Periodic flows, poisson structures, and nonholonomi mechanics. *Regul Chaotic Dyn* 10:267–284
- Fedorov YN (1999) Systems with an invariant measure on lie groups. In: *Hamiltonian systems with three or more degrees of freedom* (1995, S’Agar’o), NATO Adv Sci Inst Ser C Math Phys Sci, vol 533. Kluwer, Dordrecht, pp 350–356
- Féjóz J (2004) Démonstration du ‘théorème d’Arnold’ sur la stabilité du système planétaire (d’après Herman). *Ergod Theory Dyn Syst* 24:1521–1582. Revised version (2007) at <http://people.math.jussieu.fr/fejoz/articles.html>
- Fomenko AT (1988) *Symplectic geometry*. Gordon and Breach, New York
- Gordon WB (1969/1970) On the relation between period and energy in periodic dynamical systems. *J Math Mech* 19:111–114
- Guzzo M (1999) Nekhoroshev stability of quasi-integrable degenerate Hamiltonian systems. *Regul Chaotic Dyn* 4:78–102
- Guzzo M (2007) An overview on the Nekhoroshev theorem, Topics in gravitational dynamics, 1–28, Lecture notes in Phys. 729. Springer, Berlin
- Hanßmann H (1997) Quasi-periodic motions of a rigid body. I: quadratic Hamiltonians on the sphere with a distinguished parameter. *Regul Chaotic Dyn* 2:41–57
- Hanßmann H (2015) Perturbations of superintegrable systems. *Acta Appl Math* 137:79–95
- Herman MR (1989) Inégalités a priori pour des tores Lagrangiens invariants par des difféomorphismes symplectiques. *Inst Hautes Etudes Sci Publ Math* 70:47–101
- Holm DD, Vizman C (2012) Dual pairs in resonances. *J Geom Mech* 4:297–311
- Karasev MV, Maslov VP (1993) *Nonlinear poisson brackets. Geometry and quantization*, Translations of the AMS, vol 119. AMS, Providence
- Kibler M, Winternitz P (1990) Periodicity and quasi-periodicity for super-integrable Hamiltonian systems. *Phys Lett A* 147:338–342
- Kozlov VV (1974) Geometry of “action-angle” variables in the Euler-Poincot problem. *Vestnik Moskov Univ Ser I Mat Meh* 29:74–79
- Liebermann P, Marle C-M (1987) *Symplectic geometry and analytical mechanics*. D. Reidel, Dordrecht
- Lochak P (1992) Canonical perturbation theory via simultaneous approximation. *Russ Math Surv* 47:57–133
- Marsden JE, Ratiu TS (1994) *Introduction to mechanics and symmetry*, Texts in applied mathematics 17. Springer, New York
- Mazzocco M (1997) KAM theory for generic analytic perturbations of the Euler system. *Z Angew Math Phys* 48:193–219
- Miller W Jr, Post S, Winternitz P (2013) Classical and quantum superintegrability with applications. *J Phys A Math Theor* 46:423001 (97pp)
- Mischenko AS, Fomenko AT (1978) Generalized Liouville method of integration of Hamiltonian systems. *Funct Anal Appl* 12:113–121
- Moser J (1970) Regularization of Kepler’s problem and the averaging method on a manifold. *Commun Pure Appl Math* 23:609–636
- Neishtadt AI (1987) On the change in the adiabatic invariant on crossing a separatrix in systems with two degrees of freedom. *J Appl Math Mech* 51:586–592
- Neishtadt AI, Sidorenko VV (2004) Wisdom system: dynamics in the adiabatic approximation. *Celestial Mech Dyn Astr* 90:307–330
- Nekhoroshev NN (1972) Action-angle variables and their generalizations. *Trans Moscow Math Soc* 26:180–198
- Nekhoroshev NN (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russ Math Surv* 32:1–65
- Niederman L (1996) Stability over exponentially long times in the planetary problem. *Nonlinearity* 9:1703–1751
- Niederman L (1998) Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system. *Nonlinearity* 11:1465–1479
- Niederman L Nekhoroshev theory, *Encyclopedia of complexity and systems science*. Springer
- Ortega J-P, Ratiu TS (2004) *Momentum maps and Hamiltonian reduction*, Progress in mathematics 222. Birkhäuser Boston, Boston
- Parasyuk IO (1984) Preservation of multidimensional invariant tori of Hamiltonian systems. *Ukrain Mat Zh* 36:467–473
- Parasyuk IO (1993) Variables of the action-angle type on symplectic manifolds stratified by coisotropic tori. *Ukr Math J* 45:85–93

- Ramond P, Perez J (2021) New methods of isochrone mechanics. *J Math Phys* 62:112704
- Robutel P (1995) Stability of the planetary three-body problem. II. KAM theory and existence of quasiperiodic motions. *Celest Mech Dyn Astron* 62:219–261. See also: Erratum (2002) *Celest Mech Dyn Astron* 84:317
- Sevryuk MB (2003) The classical KAM theory at the dawn of the twenty-first century. *Mosc Math J* 3:1113–1144, 1201–1202
- Tempesta P, Winternitz P, Harnad J, Miller W, Pogosyan G, Rodriguez M (2004) Superintegrability in classical and quantum systems, CRM proceedings & lecture notes 37. AMS, Providence
- Woodhouse NMJ (1994) Geometric quantization, 2nd edn. Clarendon Press, Oxford
- Zung NT (2006) Torus actions and integrable systems. In: Bolsinov AV, Fomenko AT, Oshemkov AA (eds) *Topological methods in the theory of integrable systems*. Cambridge Scientific Publications, Cambridge, pp 289–328





# Perturbation Theory in Celestial Mechanics

Alessandra Celletti

Dipartimento di Matematica, Università di Roma  
Tor Vergata, Rome, Italy

## Article Outline

Glossary

Definition of the Subject

Introduction

Classical Perturbation Theory

Resonant Perturbation Theory

Invariant Tori

Periodic Orbits

Future Directions

Bibliography

## Glossary

**KAM Theory** Provides the persistence of quasi-periodic motions under a small perturbation of an integrable system. KAM theory can be applied under quite general assumptions, i.e., a nondegeneracy of the integrable system and a diophantine condition of the frequency of motion. It yields a constructive algorithm to evaluate the strength of the perturbation ensuring the existence of invariant tori.

**Perturbation Theory** Provides an approximate solution of the equations of motion of a nearly integrable system.

**Spin-Orbit Problem** A model composed of a rigid satellite rotating about an internal axis and orbiting around a central point-mass planet; a spin-orbit resonance means that the ratio between the revolutionary and rotational periods is rational.

**Three-Body Problem** A system composed by three celestial bodies (e.g., Sun-planet-satellite) assumed to be point-masses subject

to the mutual gravitational attraction. The restricted three-body problem assumes that the mass of one of the bodies is so small that it can be neglected.

## Definition of the Subject

Perturbation theory aims to find an approximate solution of nearly integrable systems, namely systems which are composed by an integrable part and by a small perturbation. The key point of perturbation theory is the construction of a suitable canonical transformation which removes the perturbation to higher orders. A typical example of a nearly integrable system is provided by a two-body model perturbed by the gravitational influence of a third body whose mass is much smaller than the mass of the central body. Indeed, the solution of the three-body problem greatly stimulated the development of perturbation theories. The solar system dynamics has always been a testing ground for such theories, whose applications range from the computation of the ephemerides of natural bodies to the development of the trajectories of artificial satellites.

## Introduction

The two-body problem can be solved by means of Kepler's laws, according to which for negative energies the point-mass planets move on ellipses with the Sun located in one of the two foci. The dynamics becomes extremely complicated when adding the gravitational influence of another body. Indeed Poincaré showed (Poincaré 1892) that the three-body problem does not admit a sufficient number of prime integrals which allow to integrate the problem. Nevertheless, the so-called *restricted* three-body problem deserves special attention, namely when the mass of one of the three bodies is so small that its influence on the others can be neglected. In this case, one can

assume that the primaries move on Keplerian ellipses around their common barycenter; if the mass of one of the primaries is much larger than the other (as it is the case in any Sun-planet sample), the motion of the minor body is governed by nearly integrable equations, where the integrable part represents the interaction with the major body, while the perturbation is due to the influence of the other primary. A typical example is provided by the motion of an asteroid under the gravitational attraction of the Sun and Jupiter. The small body may be taken not to influence the motion of the primaries, which are assumed to move on elliptic trajectories. The dynamics of the asteroid is essentially driven by the Sun and perturbed by Jupiter, since the Jupiter-Sun mass ratio amounts to about  $10^{-3}$ . The solution of this kind of problem stimulated the work of many scientists, especially in the eighteenth and nineteenth centuries. Indeed, Lagrange, Laplace, Leverrier, Delaunay, Tisserand, and Poincaré developed perturbation theories which are the basis of the studies of the dynamics of celestial bodies, from the computation of the ephemerides to the recent advances in flight dynamics. For example, on the basis of perturbation theory Delaunay (1867) developed a theory of the Moon, providing very refined ephemerides. Celestial Mechanics greatly motivated the advances of perturbation theories as witnessed by the discovery of Neptune: Its position was theoretically predicted by John Adams and by Jean Urbain Leverrier on the basis of perturbative computations; following the suggestion provided by the theoretical investigations, Neptune was finally discovered on 23 September 1846 by the astronomer Johann Gottfried Galle.

The aim of perturbation theory is to implement a canonical transformation which allows one to find the solution of a nearly integrable system within a better degree of approximation (see section “Classical Perturbation Theory” and references (Arnold 1978; Boccaletti and Pucacco 2001; Ferraz-Mello 2007; Hagihara 1970; Meyer and Hall 1991; Sanders and Verhulst 1985; Siegel and Moser 1971)). Let us denote the frequency vector of the system by  $\underline{\omega}$  (see “Normal Forms in Perturbation Theory,” “Kolmogorov-Arnol’d-

Moser (KAM Theory)”), which we assume to belong to  $\mathbf{R}^n$ , where  $n$  is the number of degrees of freedom of the system. Classical perturbation theory can be implemented provided that the frequency vector satisfies a nonresonant relation, which means that there does not exist a vector  $\underline{m} \in \mathbf{Z}^n$  such that  $\underline{\omega} \cdot \underline{m} \equiv \sum_{j=1}^n \omega_j m_j = 0$ . In case there exists such commensurability condition, a resonant perturbation theory can be developed as outlined in section “Resonant Perturbation Theory.” In general, the three-body problem (and, more extensively, the  $N$ -body problem) is described by a *degenerate* Hamiltonian system, which means that the integrable part (i.e., the Keplerian approximation) depends on a subset of the action variables. In this case, a degenerate perturbation theory must be implemented as explained in subsection “Degenerate Perturbation Theory.” For all the above perturbation theories (classical, resonant, and degenerate), an application to Celestial Mechanics is given: the precession of the perihelion of Mercury, orbital resonances within a three-body framework, and the precession of the equinoxes.

Even if the nonresonance condition is satisfied, the quantity  $\underline{\omega} \cdot \underline{m}$  can become arbitrarily small, giving rise to the so-called *small divisor* problem; indeed, these terms appear in the denominator of the series defining the canonical transformations necessary to implement perturbation theory, and therefore they might prevent the convergence of the series. In order to overcome the small divisor problem, a breakthrough came with the work of Kolmogorov (Kolmogorov 1954) and was later extended to different mathematical settings by Arnold (1963) and Moser (1962). The overall theory is known as the acronym KAM theory. As far as concrete estimates on the allowed size of the perturbation are concerned, the original versions of the theory gave discouraging results, which were extremely far from the physical measurements of the parameters involved in the proof. Nevertheless, the implementation of computer-assisted KAM proofs allowed one to obtain results which are in good agreement with reality. Concrete estimates with applications to Celestial Mechanics are reported in section “Invariant Tori.”

In the framework of nearly integrable systems, a very important role is provided by periodic orbits, which might be used to approximate the dynamics of quasi-periodic trajectories, for example, a truncation of the continued fraction expansion of an irrational frequency provides a sequence of rational numbers, which are associated to periodic orbits eventually approximating a quasi-periodic torus. A classical computation of periodic orbits using a perturbative approach is provided in section “Periodic Orbits,” where an application to the determination of the libration in longitude of the Moon is reported.

## Classical Perturbation Theory

### The Classical Theory

Consider a nearly integrable Hamiltonian function of the form

$$H(\underline{L}, \underline{\varphi}) = h(\underline{L}) + \varepsilon f(\underline{L}, \underline{\varphi}), \quad (1)$$

where  $h$  and  $f$  are analytic functions of  $\underline{L} \in V$  ( $V$  is an open set of  $\mathbf{R}^n$ ) and  $\underline{\varphi} \in \mathbf{T}^n$  ( $\mathbf{T}^n$  is the standard  $n$ -dimensional torus), while  $\varepsilon > 0$  is a small parameter which measures the strength of the perturbation. The aim of perturbation theory is to construct a canonical transformation, which allows to remove the perturbation to higher orders in the perturbing parameter. To this end, let us look for a canonical change of variables (i.e., with symplectic Jacobian matrix)  $C : (\underline{L}, \underline{\varphi}) \rightarrow (\underline{L}', \underline{\varphi}')$ , such that the Hamiltonian (1) takes the form

$$H'(\underline{L}', \underline{\varphi}') = H \circ C(\underline{L}, \underline{\varphi}) \equiv h'(\underline{L}') + \varepsilon^2 f'(\underline{L}', \underline{\varphi}'), \quad (2)$$

where  $h'$  and  $f'$  denote the new unperturbed Hamiltonian and the new perturbing function, respectively. To achieve such a result, we need to proceed along the following steps: Build a suitable canonical transformation close to the identity, perform a Taylor series expansion in the perturbing parameter, require that the unknown

transformation removes the dependence on the angle variables up to second-order terms, and expand in a Fourier series in order to get an explicit form of the canonical transformation.

The change of variables is defined by the equations

$$\begin{aligned} \underline{L} &= \underline{L}' + \varepsilon \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{\varphi}} \\ \underline{\varphi}' &= \underline{\varphi} + \varepsilon \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{L}'}, \end{aligned} \quad (3)$$

where  $\Phi(\underline{L}', \underline{\varphi})$  is an unknown generating function, which is determined so that (1) takes the form (2). Decompose the perturbing function as

$$f(\underline{L}, \underline{\varphi}) = f_0(\underline{L}) + \tilde{f}(\underline{L}, \underline{\varphi}),$$

where  $f_0$  is the average over the angle variables and  $\tilde{f}$  is the remainder function defined through  $\tilde{f}(\underline{L}, \underline{\varphi}) \equiv f(\underline{L}, \underline{\varphi}) - f_0(\underline{L})$ . Define the *frequency vector*  $\underline{\omega} = \underline{\omega}(\underline{L})$  as

$$\underline{\omega}(\underline{L}) \equiv \frac{\partial h(\underline{L})}{\partial \underline{L}}.$$

Inserting the transformation (3) in (1) and expanding in a Taylor series around  $\varepsilon = 0$  up to the second order, one gets

$$\begin{aligned} &h\left(\underline{L}' + \varepsilon \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{\varphi}}\right) \\ &+ \varepsilon f\left(\underline{L}' + \varepsilon \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{\varphi}}, \underline{\varphi}\right) \\ &= h(\underline{L}') + \underline{\omega}(\underline{L}') \cdot \varepsilon \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{\varphi}} + \varepsilon f_0(\underline{L}') \\ &+ \varepsilon \tilde{f}(\underline{L}', \underline{\varphi}) + O(\varepsilon^2). \end{aligned}$$

The new Hamiltonian is integrable up to  $O(\varepsilon^2)$  provided that the function  $\Phi$  satisfies:

$$\underline{\omega}(\underline{L}') \cdot \frac{\partial \Phi(\underline{L}', \underline{\varphi})}{\partial \underline{\varphi}} + \tilde{f}(\underline{L}', \underline{\varphi}) = 0. \quad (4)$$

In such case, the new integrable part becomes

$$h'(\underline{L}') = h(\underline{L}') + \varepsilon f_0(\underline{L}'),$$

which provides a better integrable approximation with respect to (1). The solution of (4) yields the explicit expression of the generating function. In fact, let us expand  $\Phi$  and  $\tilde{f}$  in Fourier series as

$$\begin{aligned} \Phi(\underline{L}', \underline{\varphi}) &= \sum_{\underline{m} \in \mathbb{Z}^n \setminus \{0\}} \hat{\Phi}_{\underline{m}}(\underline{L}') e^{i \underline{m} \cdot \underline{\varphi}}, \\ \tilde{f}(\underline{L}', \underline{\varphi}) &= \sum_{\underline{m} \in \mathcal{I}} \hat{f}_{\underline{m}}(\underline{L}') e^{i \underline{m} \cdot \underline{\varphi}}, \end{aligned} \quad (5)$$

where  $\mathcal{I}$  denotes the set of integer vectors corresponding to the nonvanishing Fourier coefficients of  $\tilde{f}$ . Inserting the above expansions in (4), one obtains

$$i \sum_{\underline{m} \in \mathbb{Z}^n \setminus \{0\}} \underline{\omega}(\underline{L}') \cdot \underline{m} \hat{\Phi}_{\underline{m}}(\underline{L}') e^{i \underline{m} \cdot \underline{\varphi}} = - \sum_{\underline{m} \in \mathcal{I}} \hat{f}_{\underline{m}}(\underline{L}') e^{i \underline{m} \cdot \underline{\varphi}},$$

which provides

$$\hat{\Phi}_{\underline{m}}(\underline{L}') = - \frac{\hat{f}_{\underline{m}}(\underline{L}')}{i \underline{\omega}(\underline{L}') \cdot \underline{m}}.$$

Casting together the above formule, the generating function is given by

$$\Phi(\underline{L}', \underline{\varphi}) = i \sum_{\underline{m} \in \mathcal{I}} \frac{\hat{f}_{\underline{m}}(\underline{L}')}{\underline{\omega}(\underline{L}') \cdot \underline{m}} e^{i \underline{m} \cdot \underline{\varphi}}. \quad (6)$$

We stress that this algorithm is constructive in the sense that it provides an explicit expression for the generating function and for the transformed Hamiltonian. We remark that (6) is well defined unless there exists an integer vector  $\underline{m} \in \mathcal{I}$  such that

$$\underline{\omega}(\underline{L}') \cdot \underline{m} = 0.$$

On the contrary, if  $\underline{\omega}$  is rationally independent, there are no zero divisors in (6), though

these terms can become arbitrarily small with a proper choice of the vector  $\underline{m}$ . This problem is known as the *small divisor problem*, which can prevent the implementation of perturbation theory (see “Normal Forms in Perturbation Theory,” “Kolmogorov-Arnol’d-Moser (KAM Theory),” and “Perturbation Theory”).

### The Precession of the Perihelion of Mercury

As an example of the implementation of classical perturbation theory, we consider the computation of the precession of the perihelion in a (restricted, planar, and circular) three-body model, taking as a sample the planet Mercury. The computation requires the introduction of Delaunay action-angle variables, the definition of the three-body Hamiltonian, the expansion of the perturbing function, and the implementation of classical perturbation theory (see Brouwer and Clemence 1961; Szebehely 1967).

### Delaunay Action: Angle Variables

We consider two bodies, say  $P_0$  and  $P_1$  with masses, respectively,  $m_0, m_1$ ; let  $M \equiv m_0 + m_1$  and let  $\mu > 0$  be a positive parameter. Let  $r$  be the orbital radius and  $\varphi$  be the longitude of  $P_1$  with respect to  $P_0$ ; let  $(I_r, I_\varphi)$  be the momenta conjugated to  $(r, \varphi)$ . In these coordinates, the two-body problem Hamiltonian takes the form

$$H_{2b}(I_r, I_\varphi, r, \varphi) = \frac{1}{2\mu} \left( I_r^2 + \frac{I_\varphi^2}{r^2} \right) - \frac{\mu M}{r}. \quad (7)$$

On the orbital plane, we introduce the planar Delaunay action-angle variables  $(\Lambda, \Gamma, \lambda, \gamma)$  as follows (Celletti and Chierchia 2007). Let  $E$  denote the total mechanical energy; then:

$$I_r = \sqrt{2\mu E + \frac{2\mu^2 M}{r} - \frac{I_\varphi^2}{r^2}}.$$

Since (7) does not depend on  $\varphi$ , setting  $\Gamma = I_\varphi$  and  $\Lambda = \sqrt{-(\mu^3 M^2)/(2E)}$ , we introduce a generating function of the form

$$F(\Lambda, \Gamma, r, \varphi) = \int \sqrt{-\frac{\mu^4 M^2}{\Lambda^2} + \frac{2\mu^2 M}{r} - \frac{\Gamma^2}{r^2}} dr + \Gamma \varphi.$$

From the definition of  $\Lambda$  the new Hamiltonian  $H_{2D}$  becomes

$$H_{2D}(\Lambda, \Gamma, \lambda, \gamma) = -\frac{\mu^3 M^2}{2\Lambda^2},$$

where  $(\Lambda, \Gamma)$  are the Delaunay action variables; by Kepler's laws, one finds that  $(\Lambda, \Gamma)$  are related to the semimajor axis  $a$  and to the eccentricity  $e$  of the Keplerian orbit of  $P_1$  around  $P_0$  by the formula:

$$\Lambda = \mu \sqrt{Ma}, \quad \Gamma = \Lambda \sqrt{1 - e^2}.$$

Concerning the conjugated angle variables, we start by introducing the eccentric anomaly  $u$  as follows: build the auxiliary circle of the ellipse, draw the line through  $P_1$  perpendicular to the semimajor axis whose intersection with the auxiliary circle forms at the origin an angle  $u$  with the semimajor axis. By the definition of the generating function, one finds

$$\begin{aligned} \lambda &= \frac{\partial F}{\partial \Lambda} = \int \frac{\mu^4 M^2}{\Lambda^3 \sqrt{-\frac{\mu^4 M^2}{\Lambda^2} + \frac{2\mu^2 M}{r} - \frac{\Gamma^2}{r^2}}} dr \\ &= u - e \sin u, \end{aligned}$$

which defines the mean anomaly  $\lambda$  in terms of the eccentric anomaly  $u$ .

In a similar way, if  $f$  denotes the true anomaly related to the eccentric anomaly by  $\tan f/2 = \sqrt{(1+e)/(1-e)} \tan u/2$ , then one has:

$$\begin{aligned} \gamma &= \frac{\partial F}{\partial \Gamma} = \varphi - \int \frac{\Gamma}{r^2 \sqrt{-\frac{\mu^4 M^2}{\Lambda^2} + \frac{2\mu^2 M}{r} - \frac{\Gamma^2}{r^2}}} dr \\ &= \varphi - f, \end{aligned}$$

which represents the argument of the perihelion of  $P_1$ , i.e., the angle between the perihelion line and a fixed reference line.

### The Restricted, Planar, Circular, Three-Body Problem

Let  $P_0$ ,  $P_1$ , and  $P_2$  be three bodies with masses  $m_0$ ,  $m_1$ , and  $m_2$ , respectively. We assume that  $m_1$  is much smaller than  $m_0$  and  $m_2$  (restricted problem) and that the motion of  $P_2$  around  $P_0$  is circular. We also assume that the three bodies always move on the same plane. We choose the free parameter  $\mu$  as  $\mu \equiv 1/m_0^{2/3}$ , so that the two-body Hamiltonian becomes  $H_{2D} = -1/(2\Lambda^2)$ , while we introduce the perturbing parameter as  $\varepsilon \equiv m_2/m_0^{2/3}$  (Celletti and Chierchia 2007). Set the units of measure so that the distance between  $P_0$  and  $P_2$  is one and so that  $m_0 + m_2 = 1$ . Taking into account the interaction of  $P_2$  on  $P_1$ , the Hamiltonian function governing the three-body problem becomes

$$\begin{aligned} H_{3b}(\Lambda, \Gamma, \lambda, \gamma, t) &= -\frac{1}{2\Lambda^2} \\ &+ \varepsilon \left( r_1 \cos(\varphi - t) - \frac{1}{\sqrt{1 + r_1^2 - 2r_1 \cos(\varphi - t)}} \right), \end{aligned}$$

where  $r_1$  is the distance between  $P_0$  and  $P_1$ . The first term of the perturbation comes out from the choice of the reference frame, while the second term is due to the interaction with the external body. Since  $\varphi - t = f + \gamma - t$ , we perform the canonical change of variables

$$\begin{aligned} L &= \Lambda \quad \ell = \lambda \\ G &= \Gamma \quad g = \gamma - t, \end{aligned}$$

which provides the following two degrees-of-freedom Hamiltonian

$$\begin{aligned} H_{3D}(L, G, \ell, g) &= -\frac{1}{2L^2} - G \\ &+ \varepsilon R(L, G, \ell, g), \end{aligned} \quad (8)$$

where

$$\begin{aligned} R(L, G, \ell, g) &\equiv r_1 \cos(\varphi - t) - \frac{1}{\sqrt{1 + r_1^2 - 2r_1 \cos(\varphi - t)}} \end{aligned} \quad (9)$$

where  $r_1$  and  $\varphi - t$  must be expressed in terms of the Delaunay variables  $(L, G, \ell, g)$ . Notice that when  $\varepsilon = 0$ , one obtains the integrable Hamiltonian function  $h(L, G) \equiv -1/(2L^2) - G$  with associated frequency vector  $\underline{\omega} = (\partial h/\partial L, \partial h/\partial G) = (1/L^3, -1)$ .

### Expansion of the Perturbing Function

We expand the perturbing function (9) in terms of the Legendre polynomials  $P_j$  obtaining

$$R = -\frac{1}{r_1} \sum_{j=2}^{\infty} P_j(\cos(\varphi - t)) \frac{1}{r_1^j}.$$

The explicit expressions of the first few Legendre polynomials are:

$$\begin{aligned} P_2(\cos(\varphi - t)) &= \frac{1}{4} + \frac{3}{4} \cos 2(\varphi - t) \\ P_3(\cos(\varphi - t)) &= \frac{3}{8} \cos(\varphi - t) + \frac{5}{8} \cos 3(\varphi - t) \\ P_4(\cos(\varphi - t)) &= \frac{9}{64} + \frac{5}{16} \cos 2(\varphi - t) \\ &\quad + \frac{35}{64} \cos 4(\varphi - t) \\ P_5(\cos(\varphi - t)) &= \frac{15}{64} \cos(\varphi - t) + \frac{35}{128} \cos 3(\varphi - t) \\ &\quad + \frac{63}{128} \cos 5(\varphi - t). \end{aligned}$$

We invert Kepler's equation  $\ell = u - e \sin u$  to the second order in the eccentricity as

$$u = \ell + e \sin \ell + \frac{e^2}{2} \sin(2\ell) + O(e^3),$$

from which one gets

$$\begin{aligned} \varphi - t &= g + \ell + 2e \sin \ell + \frac{5}{4} e^2 \sin 2\ell + O(e^3) \\ r_1 &= a \left( 1 + \frac{1}{2} e^2 - e \cos \ell - \frac{1}{2} e^2 \cos 2\ell \right) + O(e^3). \end{aligned}$$

Then, up to inessential constants the perturbing function can be expanded as

$$\begin{aligned} R &= R_{00}(L, G) + R_{10}(L, G) \cos \ell \\ &\quad + R_{11}(L, G) \cos(\ell + g) + R_{12}(L, G) \cos(\ell + 2g) \\ &\quad + R_{22}(L, G) \cos(2\ell + 2g) \\ &\quad + R_{32}(L, G) \cos(3\ell + 2g) \\ &\quad + R_{33}(L, G) \cos(3\ell + 3g) \\ &\quad + R_{44}(L, G) \cos(4\ell + 4g) \\ &\quad + R_{55}(L, G) \cos(5\ell + 5g) + \dots, \end{aligned} \quad (10)$$

where the coefficients  $R_{ij}$  are given by the following expressions (recall that  $e = \sqrt{1 - G^2/L^2}$ ):

$$\begin{aligned} R_{00} &= -\frac{L^4}{4} \left( 1 + \frac{9}{16} L^4 + \frac{3}{2} e^2 \right) + \dots, \\ R_{10} &= \frac{L^4 e}{2} \left( 1 + \frac{9}{8} L^4 \right) + \dots \\ R_{11} &= -\frac{3}{8} L^6 \left( 1 + \frac{5}{8} L^4 \right) + \dots, \\ R_{12} &= \frac{L^4 e}{4} (9 + 5L^4) + \dots \\ R_{22} &= -\frac{L^4}{4} \left( 3 + \frac{5}{4} L^4 \right) + \dots, \\ R_{32} &= -\frac{3}{4} L^4 e + \dots \\ R_{33} &= -\frac{5}{8} L^6 \left( 1 + \frac{7}{16} L^4 \right) + \dots, \\ R_{44} &= -\frac{35}{64} L^8 + \dots \\ R_{55} &= -\frac{63}{128} L^{10} + \dots \end{aligned} \quad (11)$$

### Computation of the Precession of the Perihelion

We identify the three bodies  $P_0$ ,  $P_1$ , and  $P_2$  with the Sun, Mercury, and Jupiter, respectively. Taking  $\varepsilon$  as perturbing parameter, we implement a first-order perturbation theory, which provides a new integrable Hamiltonian function of the form



$$h'(L', G') = -\frac{1}{2L'^2} - G' + \varepsilon R_{00}(L', G').$$

From Hamilton's equations, one obtains

$$\dot{g} = \frac{\partial h'(L', G')}{\partial G'} = -1 + \varepsilon \frac{\partial R_{00}(L', G')}{\partial G'};$$

neglecting  $O(\varepsilon^3)$  in  $R_{00}$  and recalling that  $g = \gamma - t$ , one has

$$\dot{\gamma} = \varepsilon \frac{\partial R_{00}(L', G')}{\partial G'} = \frac{3}{4} \varepsilon L'^2 G'.$$

Notice that to the first order in  $\varepsilon$ , one has  $L' = L$ ,  $G' = G$ . The astronomical data are  $m_0 = 2 \cdot 10^{30}$  kg,  $m_2 = 1.9 \cdot 10^{27}$  kg, which give  $\varepsilon = 9.49 \cdot 10^{-4}$ ; setting to one the Jupiter-Sun distance one has  $a = 0.0744$ ; set  $e = 0.2056$ . Taking into account that the orbital period of Jupiter amounts to about 11.86 years, one obtains

$$\dot{\gamma} = 154.65 \frac{\text{arcsecond}}{\text{century}},$$

which represents the contribution due to Jupiter to the precession of perihelion of Mercury. The value found by Leverrier on the basis of the data available in the year 1856 was of 152.59 arcsecond/century (Chebotarev 1967).

## Resonant Perturbation Theory

### The Resonant Theory

Let us consider a Hamiltonian system with  $n$  degrees of freedom of the form

$$H(\underline{L}, \underline{\varphi}) = h(\underline{L}) + \varepsilon f(\underline{L}, \underline{\varphi})$$

and let  $\omega_j(\underline{L}) = (\partial h(\underline{L}) / (\partial I_j))$  ( $j = 1, \dots, n$ ) be the frequencies of the motion, which we assume to satisfy  $\ell$ ,  $\ell < n$ , resonance relations of the form

$$\underline{\omega} \cdot \underline{m}_k = 0 \quad \text{for } k = 1, \dots, \ell,$$

for suitable rational independent integer vectors  $\underline{m}_1, \dots, \underline{m}_\ell$ . A resonant perturbation theory can be implemented to eliminate the nonresonant terms. More precisely, the aim is to construct a canonical transformation  $C : (\underline{L}, \underline{\varphi}) \rightarrow (\underline{J}', \underline{\vartheta}')$  such that the transformed Hamiltonian takes the form

$$H'(\underline{J}', \underline{\vartheta}') = h'(\underline{J}', \vartheta'_1, \dots, \vartheta'_\ell) + \varepsilon^2 f'(\underline{J}', \underline{\vartheta}'), \quad (12)$$

where  $h'$  depends only on the resonant angles  $\vartheta'_1, \dots, \vartheta'_\ell$ . To this end, let us first introduce the angles  $\underline{\vartheta} \in \mathbf{T}^n$  as

$$\begin{aligned} \vartheta_j &= \underline{m}_j \cdot \underline{\varphi} & j &= 1, \dots, \ell \\ \vartheta_k &= \underline{m}_k \cdot \underline{\varphi} & k &= \ell + 1, \dots, n, \end{aligned}$$

where the first  $\ell$  angle variables are the resonant angles, while the latter  $n - \ell$  angle variables are defined as suitable linear combinations so to make the transformation canonical together with the following change of coordinates on the actions  $\underline{J} \in \mathbf{R}^n$ :

$$\begin{aligned} I_j &= \underline{m}_j \cdot \underline{J} & j &= 1, \dots, \ell \\ I_k &= \underline{m}_k \cdot \underline{J} & k &= \ell + 1, \dots, n. \end{aligned}$$

The aim is to construct a canonical transformation which removes (to higher order) the dependence on the short-period angles  $(\vartheta_{\ell+1}, \dots, \vartheta_n)$ , while the lowest-order Hamiltonian will necessarily depend upon the resonant angles. Let us decompose the perturbation as

$$f(\underline{J}, \underline{\vartheta}) = \bar{f}(\underline{J}) + f_r(\underline{J}, \vartheta_1, \dots, \vartheta_\ell) + f_n(\underline{J}, \underline{\vartheta}), \quad (13)$$

where  $\bar{f}$  is the average of the perturbation over the angles,  $f_r$  is the part depending on the resonant angles, and  $f_n$  is the nonresonant part. In analogy to the classical perturbation theory, we implement a canonical transformation of the form

$$\underline{J} = \underline{J}' + \varepsilon \frac{\partial \Phi}{\partial \underline{\vartheta}}(\underline{J}', \underline{\vartheta})$$

$$\underline{\vartheta}' = \underline{\vartheta} + \varepsilon \frac{\partial \Phi}{\partial \underline{J}'}(\underline{J}', \underline{\vartheta}),$$

such that the new Hamiltonian takes the form (12). Taking into account (13) and developing up to the second order in the perturbing parameter, one obtains:

$$\begin{aligned} & h\left(\underline{J}' + \varepsilon \frac{\partial \Phi}{\partial \underline{\vartheta}}\right) + \varepsilon f(\underline{J}', \underline{\vartheta}) + O(\varepsilon^2) \\ &= h(\underline{J}') + \varepsilon \sum_{k=1}^n \frac{\partial h}{\partial J_k} \frac{\partial \Phi}{\partial \vartheta_k} + \varepsilon \bar{f}(\underline{J}') \\ &+ \varepsilon f_r(\underline{J}', \vartheta_1, \dots, \vartheta_\ell) + \varepsilon f_n(\underline{J}', \underline{\vartheta}) \\ &+ O(\varepsilon^2). \end{aligned}$$

Equating same orders of  $\varepsilon$ , one gets that

$$\begin{aligned} h'(\underline{J}', \vartheta'_1, \dots, \vartheta'_\ell) &= h(\underline{J}') + \varepsilon \bar{f}(\underline{J}') \\ &+ \varepsilon f_r(\underline{J}', \vartheta'_1, \dots, \vartheta'_\ell), \end{aligned} \quad (14)$$

provided that

$$\sum_{k=1}^n \omega'_k \frac{\partial \Phi}{\partial \vartheta_k} = -f_n(\underline{J}', \underline{\vartheta}), \quad (15)$$

where  $\omega'_k = \omega'_k(\underline{J}') \equiv (\partial h(\underline{J}') / \partial J'_k)$ . The solution of (15) gives the generating function, which allows one to reduce the Hamiltonian to the required form (12); as a consequence, the conjugated action variables, say  $J_{\ell+1}', \dots, J'_n$ , are constants of the motion up to the second order in  $\varepsilon$ . We conclude by mentioning that using the new frequencies  $\omega'_k$ , the resonant relations take the form  $\omega'_k = 0$  for  $k = 1, \dots, \ell$ .

### Three-Body Resonance

We consider the three-body Hamiltonian (8) with perturbing function (10)–(11) and let  $\underline{\omega} \equiv (\omega_\ell, \omega_g)$  be the frequency of motion. We assume that the frequency vector satisfies the resonance relation

$$\omega_\ell + 2\omega_g = 0.$$

According to the theory described in the previous section, we perform the canonical change of variables

$$\begin{aligned} \vartheta_1 &= \ell + 2g & J_1 &= \frac{1}{2}G \\ \vartheta_2 &= 2\ell & J_2 &= \frac{1}{2}L - \frac{1}{4}G. \end{aligned}$$

In the new coordinates, the unperturbed Hamiltonian becomes

$$h'(\underline{J}) \equiv -\frac{1}{2(J_1 + 2J_2)^2} - 2J_1,$$

with frequency vector  $\underline{\omega}' \equiv \frac{\partial h'(\underline{J})}{\partial \underline{J}}$ , while the perturbation takes the form

$$\begin{aligned} R(J_1, J_2, \vartheta_1, \vartheta_2) &\equiv R_{00}(\underline{J}) \\ &+ R_{10}(\underline{J}) \cos\left(\frac{1}{2}\vartheta_2\right) \\ &+ R_{11}(\underline{J}) \cos\left(\frac{1}{2}\vartheta_1 + \frac{1}{4}\vartheta_2\right) \\ &+ R_{12}(\underline{J}) \cos(\vartheta_1) \\ &+ R_{22}(\underline{J}) \cos\left(\vartheta_1 + \frac{1}{2}\vartheta_2\right) \\ &+ R_{32}(\underline{J}) \cos(\vartheta_1 + \vartheta_2) \\ &+ R_{33}(\underline{J}) \cos\left(\frac{3}{2}\vartheta_1 + \frac{3}{4}\vartheta_2\right) \\ &+ R_{44}(\underline{J}) \cos(2\vartheta_1 + \vartheta_2) \\ &+ R_{55}(\underline{J}) \cos\left(\frac{5}{2}\vartheta_1 + \frac{5}{4}\vartheta_2\right) + \dots \end{aligned}$$

with the coefficients  $R_{ij}$  as in (11). Let us decompose the perturbation as  $R = \bar{R}(\underline{J}) + R_r(\underline{J}, \vartheta_1) + R_n(\underline{J}, \underline{\vartheta})$ , where  $\bar{R}(\underline{J})$  is the average over the angles,  $R_r(\underline{J}, \vartheta_1) = R_{12}(\underline{J}) \cos(\vartheta_1)$  is the resonant part, while  $R_n$  contains all the remaining nonresonant terms. We look for a canonical transformation close to the identity with generating function  $\Phi = \Phi(\underline{J}', \underline{\vartheta})$  such that

$$\underline{\omega}'(\underline{J}') \cdot \frac{\partial \Phi(\underline{J}', \underline{\vartheta})}{\partial \underline{\vartheta}} = -R_n(\underline{J}', \underline{\vartheta}),$$

which is well defined since  $\underline{\omega}'$  is nonresonant for the Fourier components appearing in  $R_n$ . Finally, according to (14) the new unperturbed Hamiltonian is given by

$$h'(\underline{J}', \vartheta'_1) \equiv h(\underline{J}') + \varepsilon R_{00}(\underline{J}') + \varepsilon R_{12}(\underline{J}') \cos \vartheta'_1.$$

### Degenerate Perturbation Theory

A special case of resonant perturbation theory is obtained when considering a degenerate Hamiltonian function with  $n$  degrees of freedom of the form

$$H(\underline{L}, \underline{\varphi}) = h(I_1, \dots, I_d) + \varepsilon f(\underline{L}, \underline{\varphi}), \quad d < n; \quad (16)$$

notice that the integrable part depends on a subset of the action variables, being degenerate in  $I_{d+1}, \dots, I_n$ . In this case, we look for a canonical transformation  $C: (\underline{L}, \underline{\varphi}) \rightarrow (\underline{L}', \underline{\varphi}')$  such that the transformed Hamiltonian becomes

$$H'(\underline{L}', \underline{\varphi}') = h'(\underline{L}') + \varepsilon h'_1(\underline{L}, \varphi'_{d+1}, \dots, \varphi'_n) + \varepsilon^2 f'(\underline{L}', \underline{\varphi}'), \quad (17)$$

where the part  $h' + \varepsilon h'_1$  admits  $d$  integrals of motion. Let us decompose the perturbing function in (16) as

$$f(\underline{L}, \underline{\varphi}) = \bar{f}(\underline{L}) + f_d(\underline{L}, \varphi_{d+1}, \dots, \varphi_n) + \tilde{f}(\underline{L}, \underline{\varphi}), \quad (18)$$

where  $\bar{f}$  is the average over the angle variables,  $f_d$  is independent on  $\varphi_1, \dots, \varphi_d$ , and  $\tilde{f}$  is the remainder. As in the previous sections, we want to determine a near-to-identity canonical transformation  $\Phi = \Phi(\underline{L}', \underline{\varphi})$  of the form (3), such that in view of (18) the Hamiltonian (16) takes the form (17). One obtains

$$\begin{aligned} & h(I'_1, \dots, I'_d) + \varepsilon \sum_{k=1}^d \frac{\partial h}{\partial I_k} \frac{\partial \Phi}{\partial \varphi_k} + \varepsilon \bar{f}(I') \\ & + \varepsilon f_d(\underline{L}', \varphi_{d+1}, \dots, \varphi_n) + \varepsilon \tilde{f}(\underline{L}', \underline{\varphi}) + O(\varepsilon^2) \\ & = h'(\underline{L}') + \varepsilon h'_1(\underline{L}', \varphi_{d+1}, \dots, \varphi_n) + O(\varepsilon^2), \end{aligned}$$

where

$$\begin{aligned} h'(\underline{L}') &= h(I'_1, \dots, I'_d) + \varepsilon \bar{f}(I') \\ h'_1(\underline{L}', \varphi_{d+1}, \dots, \varphi_n) &= f_d(\underline{L}', \varphi_{d+1}, \dots, \varphi_n), \end{aligned}$$

while  $\Phi$  is determined solving the equation

$$\sum_{k=1}^d \frac{\partial h}{\partial I_k} \frac{\partial \Phi}{\partial \varphi_k} + \tilde{f}(\underline{L}', \underline{\varphi}) = 0.$$

Expanding  $\Phi$  and  $\tilde{f}$  in Fourier series as in (5), one obtains that  $\Phi$  is given by (6) where  $\underline{\omega} \cdot \underline{m} = \sum_{k=1}^d m_k \omega_k$ , being  $\omega_k = 0$  for  $k = d+1, \dots, n$ . The generating function is well defined provided that  $\underline{\omega} \cdot \underline{m} \neq 0$  for any  $\underline{m} \in \mathcal{I}$ , which is equivalent to requiring that

$$\sum_{k=1}^d m_k \omega_k \neq 0 \quad \text{for } \underline{m} \in \mathcal{I}.$$

### The Precession of the Equinoxes

An example of the application of the degenerate perturbation theory in Celestial Mechanics is provided by the computation of the precession of the equinoxes.

We consider a triaxial rigid body moving in the gravitational field of a primary body. We introduce the following reference frames with a common origin in the barycenter of the rigid body:  $(O, i_1^{(i)}, i_2^{(i)}, i_3^{(i)})$  is an inertial reference frame,  $(O, i_1^{(b)}, i_2^{(b)}, i_3^{(b)})$  is a body frame oriented along the direction of the principal axes of the ellipsoid, and  $(O, i_1^{(s)}, i_2^{(s)}, i_3^{(s)})$  is the spin reference frame with the vertical axis along the direction of the angular momentum. Let  $(J, g, \ell)$  be the Euler angles formed by the body and spin frames, and let  $(K, h, 0)$  be the Euler angles formed by the spin and inertial frames. The angle  $K$  is the obliquity (representing the angle between the spin and inertial vertical axes), while  $J$  is the nonprincipal rotation angle (representing the angle between the spin and body vertical axes).

This problem is conveniently described in terms of the following set of action-angle variables introduced by Andoyer in (1926) (see also Deprit 1967). Let  $\underline{M}_0$  be the angular momentum and let  $M_0 \equiv |\underline{M}_0|$ ; the action variables are defined as

$$\begin{aligned} G &\equiv \underline{M}_0 \cdot \dot{\underline{l}}_3^{(s)} = M_0 \\ L &\equiv \underline{M}_0 \cdot \dot{\underline{l}}_3^{(b)} = G \cos J \\ H &\equiv \underline{M}_0 \cdot \dot{\underline{l}}_3^{(i)} = G \cos K, \end{aligned}$$

while the corresponding angle variables are the quantities  $(g, \ell, h)$  introduced before.

We limit ourselves to consider the gyroscopic case in which  $I_1 = I_2 < I_3$  are the principal moments of inertia of the rigid body  $E$  around the primary  $S$ ; let  $m_E$  and  $m_S$  be their masses, and let  $|E|$  be the volume of  $E$ . We assume that  $E$  orbits on a Keplerian ellipse around  $S$  with semimajor axis  $a$  and eccentricity  $e$ , while  $\lambda_E$  and  $r_E$  denote the longitude and instantaneous orbital radius (due to the assumption of Keplerian motion,  $\lambda_E$  and  $r_E$  are known functions of the time). The Hamiltonian describing the motion of  $E$  around  $S$  is given by Chierchia and Gallavotti (1994).

$$\begin{aligned} \mathcal{H}(L, G, H, \ell, g, h, t) \\ = \frac{G^2}{2I_1} + \frac{I_1 - I_3}{2I_1 I_3} L^2 + V(L, G, H, \ell, g, h, t), \end{aligned}$$

where the perturbation is implicitly defined by

$$V \equiv - \int_E \frac{\tilde{G} m_S m_E}{|\underline{r}_E + \underline{x}|} \frac{d\underline{x}}{|E|},$$

$\tilde{G}$  being the gravitational constant. Setting  $r_E = |\underline{r}_E|$  and  $x = |\underline{x}|$ , we can expand  $V$  using the Legendre polynomials as

$$\begin{aligned} V = - \frac{\tilde{G} m_S m_E}{r_E} \\ \int_E \frac{d\underline{x}}{|E|} \left[ 1 - \frac{\underline{x} \cdot \underline{r}_E}{r_E^2} + \frac{1}{2r_E^2} \cdot \left( 3 \frac{(\underline{x} \cdot \underline{r}_E)^2}{r_E^2} - x^2 \right) \right] \\ + O\left(\left(\frac{x}{r_E}\right)^3\right). \end{aligned}$$

We further assume that  $J = 0$  (i.e.,  $G = L$ ) so that  $E$  rotates around a principal axis. Let  $G_0$  and  $H_0$  be the initial values of  $G$  and  $H$ ; if  $\alpha$  denotes the angle between  $\underline{r}_E$  and  $\dot{\underline{l}}_3^{(b)}$ , the perturbing function can be written as

$$V = \frac{3}{2} \eta \omega \frac{G_0^2}{H_0} \frac{(1 - e \cos \lambda_E)^3}{(1 - e^2)^3} \cos^2 \alpha$$

with  $\eta = (I_3 - I_1)/I_3$  and  $\omega = (\tilde{G} m_S)/a^3 I_3 H_0 / G_0^2$ . Elementary computations show that

$$\cos \alpha = \sin(\lambda_E - h) \sqrt{1 - \frac{H^2}{G^2}}.$$

Neglecting first-order terms in the eccentricity, we approximate  $(1 - e \cos \lambda_E)^3 / (1 - e^2)^3$  with one. A first-order degenerate perturbation theory provides that the new unperturbed Hamiltonian is given by

$$\mathcal{K}(G, H) = \frac{G^2}{2I_3} + \frac{3}{2} \eta \omega \frac{G_0^2}{H_0} \frac{G^2 - H^2}{2G^2}.$$

Therefore, the average angular velocity of precession is given by

$$\dot{h} = \frac{\partial \mathcal{K}(G, H)}{\partial H} = -\frac{3}{2} \eta \omega \frac{G_0^2}{H_0} \frac{H}{G^2}.$$

At  $t = 0$ , it is

$$\dot{h} = -\frac{3}{2} \eta \omega = -\frac{3}{2} \eta \omega_y^2 \omega_d^{-1} \cos K, \quad (19)$$

where we used  $\omega = \omega_y^2 \omega_d^{-1} \cos K$  with  $\omega_y$  being the frequency of revolution and  $\omega_d$  the frequency of rotation.

In the case of the Earth, the astronomical measurements show that  $\eta = 1/298.25$ ,  $K = 23.45^\circ$ . The contribution due to the Sun is thus obtained by inserting  $\omega_y = 1$  year,  $\omega_d = 1$  day in (19), which yields  $\dot{h}^{(S)} = -2.51857 \cdot 10^{-12}$  rad/sec, corresponding to a retrograde precessional period of 79,107.9 years. A similar computation shows that the contribution of the Moon amounts to

$\dot{h}^{(L)} = -5.49028 \cdot 10^{-12}$  rad/sec, corresponding to a precessional period of 36,289.3 years. The total amount is obtained as the sum of  $\dot{h}^{(S)}$  and  $\dot{h}^{(L)}$ , providing an overall retrograde precessional period of 24,877.3 years.

## Invariant Tori

### Invariant KAM Surfaces

We consider an  $n$ -dimensional nearly integrable Hamiltonian function

$$H(\underline{I}, \underline{\varphi}) = h(\underline{I}) + \varepsilon f(\underline{I}, \underline{\varphi}),$$

defined in a  $2n$ -dimensional phase space  $\mathcal{M} \equiv V \times \mathbf{T}^n$ , where  $V$  is an open bounded region of  $\mathbf{R}^n$ . A KAM torus associated to  $H$  is an  $n$ -dimensional invariant surface on which the flow is described parametrically by a coordinate  $\underline{\theta} \in \mathbf{T}^n$  such that the *conjugated flow* is linear, namely  $\underline{\theta} \in \mathbf{T}^n \rightarrow \underline{\theta} + \underline{\omega}t$  where  $\underline{\omega} \in \mathbf{R}^n$  is a Diophantine vector, i.e., there exist  $\gamma > 0$  and  $\tau > 0$  such that

$$|\underline{\omega} \cdot \underline{m}| \geq \frac{\gamma}{|\underline{m}|^\tau}, \quad \forall \underline{m} \in \mathbf{Z}^n \setminus \{0\}.$$

Kolmogorov's theorem (1954) (see also "Kolmogorov-Arnold-Moser (KAM) Theory for Finite and Infinite Dimensional Systems") ensures the persistence of invariant tori with diophantine frequency, provided  $\varepsilon$  is sufficiently small and provided the unperturbed Hamiltonian is non degenerate, i.e., for a given torus  $\{L_0\} \times \mathbf{T}^n \subset \mathcal{M}$

$$\det h''(L_0) \equiv \det \left( \frac{\partial^2 h}{\partial I_i \partial I_j} (L_0) \right)_{i,j=1,\dots,n} \neq 0. \quad (20)$$

The condition (20) can be replaced by the isoenergetic nondegeneracy condition introduced by Arnold (1963).

$$\det \begin{pmatrix} h''(\underline{I}_0) & h'(\underline{I}_0) \\ h'(\underline{I}_0) & \underline{0} \end{pmatrix} \neq 0, \quad (21)$$

which ensures the existence of KAM tori on the energy level corresponding to the unperturbed

energy  $h(L_0)$ , say  $\mathcal{M}_0 \equiv \{(\underline{I}, \underline{\varphi}) \in \mathcal{M} : H(\underline{I}, \underline{\varphi}) = h(L_0)\}$ . In the context of the  $n$ -body problem, Arnold (1963) addressed the question of the existence of a set of initial conditions with positive measure such that, if the initial position and velocities of the bodies belong to this set, then the mutual distances remain perpetually bounded. A positive answer is provided by Kolmogorov's theorem in the framework of the planar, circular, restricted three-body problem, since the integrable part of the Hamiltonian (8) satisfies the isoenergetic nondegeneracy condition (21); denoting the initial values of the Delaunay's action variables by  $(L_0, G_0)$ , if  $\varepsilon$  is sufficiently small, there exist KAM tori for (8) on the energy level  $\mathcal{M}_0 \equiv \{H_{3D} = -1/(2L_0^2) - G_0\}$ . In particular, the motion of the perturbed body remains forever bounded from the orbits of the primaries. Indeed, a stronger statement is also valid: Due to the fact that the two-dimensional KAM surfaces separate the three-dimensional energy levels, any trajectory starting between two KAM tori remains forever trapped in the region between such tori.

In the framework of the three-body problem, Arnold (1963) stated the following result: "*If the masses, eccentricities and inclinations of the planets are sufficiently small, then for the majority of initial conditions the true motion is conditionally periodic and differs little from Lagrangian motion with suitable initial conditions throughout an infinite interval time  $-\infty < t < \infty$ .*" Arnold provided a complete proof for the case of three coplanar bodies, while the spatial three-body problem was investigated by Laskar and Robutel in Laskar and Robutel (1995); Robutel (1995) using Poincaré variables, the Jacobi's "reduction of the nodes" (see Celletti and Chierchia 2006) and Birkhoff's normal form Arnold (1978), (1988); Siegel and Moser (1971). The full proof of Arnold's theorem was provided in Féjóz (2004), based on Herman's results on the planetary problem; it makes use of Poincaré variables restricted to the symplectic manifold of vertical total angular momentum.

Explicit estimates on the perturbing parameter ensuring the existence of KAM tori were given by M. Hénon (1966); he showed that direct

applications of KAM theory to the three-body problem lead to analytical results which are much smaller than the astronomical observations. For example, the application of Arnold's theorem to the restricted three-body problem is valid provided the mass ratio of the primaries is less than  $10^{-333}$ . This result can be improved up to  $10^{-48}$  by applying Moser's theorem, but it is still very far from the actual Jupiter-Sun mass ratio which amounts to about  $10^{-3}$ . In the context of concrete estimates, a big improvement comes from the synergy between KAM theory and computer-assisted proofs, based on the application of *interval arithmetic* which allows to rigorously keep track of the rounding-off and propagation errors introduced by the machine. Computer-assisted KAM estimates were implemented in a number of cases in Celestial Mechanics, like the three-body problem and the spin-orbit model as briefly recalled in the following subsections.

Another interesting example of the interaction between the analytical theory and the computer implementation is provided by the analysis of the stability of the triangular Lagrangian points; in particular, the stability for exponentially long times is obtained using Nekhoroshev theory combined with computer-assisted implementations of Birkhoff normal form (see Benettin et al. 1998; Celletti and Giorgilli 1991; Efthymiopoulos and Sandor 2005; Gabern et al. 2005; Giorgilli and Skokos 1997; Giorgilli et al. 1989; Lhotka et al. 2008; Robutel and Gabern 2006).

### Rotational Tori for the Spin-Orbit Problem

We study the motion of a rigid triaxial satellite around a central planet under the following assumptions (Celletti 1990):

1. The orbit of the satellite is Keplerian.
2. The spin-axis is perpendicular to the orbital plane.
3. The spin-axis coincides with the smallest physical axis.
4. External perturbations as well as dissipative forces are neglected.

Let  $I_1 < I_2 < I_3$  be the principal moments of inertia; let  $a$ ,  $e$  be the semimajor axis and

eccentricity of the Keplerian ellipse; let  $r$  and  $f$  be the instantaneous orbital radius and the true anomaly of the satellite; and let  $x$  be the angle between the longest axis of the triaxial satellite and the periaapsis line. The equation of motion governing the spin-orbit model is given by:

$$\ddot{x} + \frac{3}{2} \frac{I_2 - I_1}{I_3} \left(\frac{a}{r}\right)^3 \sin(2x - 2f) = 0. \quad (22)$$

Due to *assumption 1*, the quantities  $r$  and  $f$  are known functions of the time. Expanding the second term of (22) in Fourier-Taylor series and neglecting terms of order 6 in the eccentricity, setting  $y \equiv \dot{x}$  one obtains that the equation of motion corresponds to Hamilton's equations associated to the Hamiltonian

$$\begin{aligned} H(y, x, t) \equiv & \frac{y^2}{2} - \varepsilon \left[ \left( -\frac{e}{4} + \frac{e^3}{32} - \frac{5}{768}e^5 \right) \cos(2x - t) \right. \\ & + \left( \frac{1}{2} - \frac{5}{4}e^2 + \frac{13}{32}e^4 \right) \cos(2x - 2t) \\ & + \left( \frac{7}{4}e - \frac{123}{32}e^3 + \frac{489}{256}e^5 \right) \cos(2x - 3t) \\ & + \left( \frac{17}{4}e^2 - \frac{115}{12}e^4 \right) \cos(2x - 4t) \\ & + \left( \frac{845}{96}e^3 - \frac{32525}{1536}e^5 \right) \cos(2x - 5t) \\ & + \frac{533}{32}e^4 \cos(2x - 6t) \\ & \left. + \frac{228347}{7680}e^5 \cos(2x - 7t) \right], \end{aligned} \quad (23)$$

where  $\varepsilon \equiv 3/2(I_2 - I_1)/I_3$ , and we have chosen the units so that  $a = 1$ ,  $2\pi/T_{\text{rev}} = 1$ , where  $T_{\text{rev}}$  is the period of revolution. Let  $p$ ,  $q$  be integers with  $q \neq 0$ ; a  $p : q$  resonance occurs whenever  $\langle \dot{x} \rangle = \frac{p}{q}$ , meaning that during  $q$  orbital revolutions, the satellite makes on average  $p$  rotations. Since the phase-space is three-dimensional, the two-dimensional KAM tori separate the phase-space into invariant regions, thus providing the stability of the trapped orbits. In particular, let



$\mathcal{P}\left(\frac{p}{q}\right)$  be the periodic orbit associated to the  $p : q$  resonance; its stability is guaranteed by the existence of trapping rotational tori with frequencies  $\mathcal{T}(\omega_1)$  and  $\mathcal{T}(\omega_2)$  with  $\omega_1 < p/q < \omega_2$ . For example, one can consider the sequences of irrational rotation numbers

$$\begin{aligned}\Gamma_k^{(p/q)} &\equiv \frac{p}{q} - \frac{1}{k + \alpha}, \quad \Delta_k^{(p/q)} \\ &\equiv \frac{p}{q} + \frac{1}{k + \alpha}, k \in \mathbf{Z}, k \geq 2\end{aligned}$$

with  $\alpha \equiv (\sqrt{5} - 1)/2$ . In fact, the continued fraction expansion of  $1/(k + \alpha)$  is given by  $1/(k + \alpha) = [0, k, 1^\infty]$ . Therefore, both  $\Gamma_k^{(p/q)}$  and  $\Delta_k^{(p/q)}$  are *noble* numbers (i.e., with continued fraction expansion definitely equal to one); by number theory, they satisfy the diophantine condition and bound  $\frac{p}{q}$  from below and above.

As a concrete sample, we consider the synchronous spin-orbit resonance ( $p = q = 1$ ) of the Moon, whose physical values of the parameters are  $\varepsilon \equiv 3.45 \cdot 10^{-4}$  and  $e = 0.0549$ . The stability of the motion is guaranteed by the existence of the surfaces  $\mathcal{T}\left(\Gamma_{40}^{(1)}\right)$  and  $\mathcal{T}\left(\Delta_{40}^{(1)}\right)$ , which is obtained by implementing a computer-assisted KAM theory for the realistic values of the parameters. The result provides the confinement of the synchronous periodic orbit in a limited region of the phase space Celletti (1990).

### Librational Tori for the Spin-Orbit Problem

The existence of invariant librational tori around a spin-orbit resonance can be obtained as follows (Celletti 1993). Let us consider the 1:1 resonance corresponding to Hamilton's equations associated to (23). First one implements a canonical transformation to center around the synchronous periodic orbit; after expanding in Taylor series, one diagonalizes the quadratic terms, thus obtaining a harmonic oscillator plus higher degree (time-dependent) terms. Finally, it is convenient to transform the Hamiltonian using the action-angle variables  $(I, \varphi)$  of the harmonic oscillator. After these symplectic changes of variables, one is led to a Hamiltonian of the form

$$\begin{aligned}H(I, \varphi, t) &\equiv \omega I + \varepsilon \bar{h}(I) + \varepsilon R(I, \varphi, t), \\ I &\in \mathbf{R}, (\varphi, t) \in T^2,\end{aligned}$$

where  $\omega \equiv \omega(\varepsilon)$  is the frequency of the harmonic oscillator, while  $\bar{h}(I)$  and  $R(I, \varphi, t)$  are suitable functions, precisely polynomials in the action (of order of the square of the action). Then apply a Birkhoff normal form (see "Normal Forms in Perturbation Theory") up to the order  $k$  ( $k = 5$  in (Celletti 1993)) to obtain the following Hamiltonian:

$$H_k(I', \varphi', t) = \omega I' + \varepsilon h_k(I'; \varepsilon) + \varepsilon^{k+1} R_k(I', \varphi', t).$$

Finally, implementing a computer-assisted KAM theorem, one gets the following result: Consider the Moon-Earth case with  $\varepsilon_{\text{obs}} = 3.45 \cdot 10^{-4}$  and  $e = 0.0549$ ; there exists an invariant torus around the synchronous resonance corresponding to a libration of  $8.79^\circ$  for any  $\varepsilon \leq \varepsilon_{\text{obs}}/5.26$ . The same strategy applied to different samples, e.g., the Rhea-Saturn pair, allows one to prove the existence of librational invariant tori around the synchronous resonance for values of the parameters in full agreement with the observational measurements (Celletti 1993).

### Rotational Tori for the Restricted Three-Body Problem

The planar, circular, restricted three-body problem has been considered in Celletti and Chierchia (2007), where the stability of the asteroid 12 Victoria has been investigated under the gravitational influence of the Sun and Jupiter. On a fixed energy level, invariant KAM tori trapping the motion of Victoria have been established for the astronomical value of the Jupiter-Sun mass ratio (about  $10^{-3}$ ). After an expansion of the perturbing function and a truncation to a suitable order (see Celletti and Chierchia 2007), the Hamiltonian function describing the motion of the asteroid is given in Delaunay's variables by

$$H(L, G, \ell, g) \equiv -\frac{1}{2L^2} - G - \varepsilon f(L, G, \ell, g),$$

where setting  $a \equiv L^2$ ,  $e = \sqrt{1 - G^2/L^2}$ , the perturbation is given by

$$\begin{aligned}
f(L, G, \ell, g) = & 1 + \frac{a^2}{4} + \frac{9}{64}a^4 + \frac{3}{8}a^2e^2 \\
& - \left(\frac{1}{2} + \frac{9}{16}a^2\right)a^2e \cos \ell \\
& + \left(\frac{3}{8}a^3 + \frac{15}{64}a^5\right) \cos(\ell + g) \\
& - \left(\frac{9}{4} + \frac{5}{4}a^2\right)a^2e \cos(\ell + 2g) \\
& + \left(\frac{3}{4}a^2 + \frac{5}{16}a^4\right) \cos(2\ell + 2g) \\
& + \frac{3}{4}a^2e \cos(3\ell + 2g) \\
& + \left(\frac{5}{8}a^3 + \frac{35}{128}a^5\right) \cos(3\ell + 3g) \\
& + \frac{35}{64}a^4 \cos(4\ell + 4g) \\
& + \frac{63}{128}a^5 \cos(5\ell + 5g).
\end{aligned}$$

For the asteroid Victoria, the orbital elements are  $a_V \simeq 2.334$  AU,  $e_V \simeq 0.220$ , which give the observed values of the Delaunay's action variables as  $L_V = 0.670$ ,  $G_V = 0.654$ . The energy level is taken as

$$\begin{aligned}
E_V^{(0)} &\equiv -\frac{1}{2L_V^2} - G_V \simeq -1.768, \\
E_V^{(1)} &\equiv -\langle f(L_V, G_V, \ell, g) \rangle \simeq -1.060, \\
E_V(\varepsilon) &\equiv E_V^{(0)} + \varepsilon E_V^{(1)}.
\end{aligned}$$

The osculating energy level of the Sun-Jupiter-Victoria model is defined as

$$E_V^* \equiv E_V(\varepsilon_J) = E_V^{(0)} + \varepsilon_J E_V^{(1)} \simeq -1.769.$$

We now look for two invariant tori bounding the observed values of  $L_V$  and  $G_V$ . To this end, let  $\tilde{L}_\pm = L_V \pm 0.001$  and let

$$\tilde{\omega}_\pm = \left( \frac{1}{\tilde{L}_\pm^3}, -1 \right) \equiv (\tilde{\alpha}_\pm, -1).$$

To obtain diophantine frequencies, the continued fraction expansion of  $\tilde{\alpha}_\pm$  is modified adding a

tail of ones after the order 5; this procedure gives the diophantine numbers  $\alpha_\pm$  which define the bounding frequencies as  $\underline{\omega}_\pm = (\alpha_\pm, -1)$ . By a computer-assisted KAM theorem, the stability of the asteroid Victoria is a consequence of the following result (Cellesti and Chierchia 2007): For  $|\varepsilon| \leq 10^{-3}$ , the unperturbed tori can be analytically continued into invariant KAM tori for the perturbed system on the energy level  $H^{-1}(E_V(\varepsilon))$ , keeping fixed the ratio of the frequencies. Therefore, the orbital elements corresponding to the semimajor axis and to the eccentricity of the asteroid Victoria stay forever  $\varepsilon$ -close to their unperturbed values.

### Planetary Problem

The dynamics of the planetary problem composed by the Sun, Jupiter, and Saturn is investigated in Locatelli and Giorgilli (2007), (2005), (2000). In Locatelli and Giorgilli (2000), the secular dynamics of the following model is studied: After the Jacobi's reduction of the nodes, the four-dimensional Hamiltonian is averaged over the fast angles and its series expansion is considered up to the second order in the masses. This procedure provides a Hamiltonian function with two degrees of freedom, describing the slow motion of the parameters characterizing the Keplerian approximation (i.e., the eccentricities and the arguments of perihelion). Afterward, action-angle coordinates are introduced and a partial Birkhoff normalization is performed. Finally, a computer-assisted implementation of a KAM theorem yields the existence of two invariant tori bounding the secular motions of Jupiter and Saturn for the observed values of the parameters.

The approach sketched above is extended in Locatelli and Giorgilli (2007) so to include the description of the fast variables, like the semimajor axes and the mean longitudes of the planets. Indeed, the preliminary average on the fast angles is now performed without eliminating the terms with degree greater or equal than two with respect to the fast actions. The canonical transformations involving the secular coordinates can be adapted to produce a good initial approximation of an invariant torus for the reduced Hamiltonian of the

three-body planetary problem. This is the starting point of the procedure for constructing the Kolmogorov's normal form which is numerically shown to be convergent. In Locatelli and Giorgilli (2005), the same result of Locatelli and Giorgilli (2007) has been obtained for a fictitious planetary solar system composed by two planets with masses equal to 1/10 of those of Jupiter and Saturn.

## Periodic Orbits

### Construction of Periodic Orbits

One of the most intriguing conjectures of Poincaré concerns the pivotal role of the periodic orbits in the study of the dynamics; more precisely, he states that given a particular solution of Hamilton's equations, one can always find a periodic solution (possibly with very long period) such that the difference between the two solutions is small for an arbitrary long time. The literature on periodic orbits is extremely wide (see Arnold 1988; Brouwer and Clemence 1961; Hagihara 1970; Siegel and Moser 1971; Szebehely 1967) and references therein); here we present the construction of periodic orbits implementing a perturbative approach (see Celletti and Chierchia 1998) as shown by Poincaré in (1892). We describe such a method taking as an example the spin-orbit Hamiltonian (23) that we write in a compact form as  $H(y, x, t) \equiv y^2/2 - \varepsilon f(x, t)$  for a suitable function  $f = f(x, t)$ ; the corresponding Hamilton's equations are

$$\begin{aligned}\dot{x} &= y \\ \dot{y} &= \varepsilon f_x(x, t).\end{aligned}\tag{24}$$

A spin-orbit resonance of order  $p : q$  is a periodic solution of period  $T = 2\pi q$  ( $q \in \mathbf{Z} \setminus \{0\}$ ), such that

$$\begin{aligned}x(t + 2\pi q) &= x(t) + 2\pi p \\ y(t + 2\pi q) &= y(t).\end{aligned}\tag{25}$$

From (24), the solution can be written in integral form as

$$\begin{aligned}y(t) &= y(0) + \varepsilon \int_0^t f_x(x(s), s) ds \\ x(t) &= x(0) + y(0)t + \varepsilon \int_0^t \int_0^\tau f_x(x(s), s) ds d\tau \\ &= x(0) + \int_0^t y(s) ds;\end{aligned}$$

combining the above equations with (25), one obtains

$$\begin{aligned}\int_0^{2\pi q} f_x(x(s), s) ds &= 0 \\ \int_0^{2\pi q} y(s) ds - 2\pi p &= 0.\end{aligned}\tag{26}$$

Let us write the solution as the series

$$\begin{aligned}x(t) &\equiv \bar{x} + \bar{y}t + \varepsilon x_1(t) + \dots \\ y(t) &\equiv \bar{y} + \varepsilon y_1(t) + \dots,\end{aligned}\tag{27}$$

where  $x(0) = \bar{x}$  and  $y(0) = \bar{y}$  are the initial conditions, while  $x_1(t), y_1(t)$  are the first-order terms in  $\varepsilon$ . Expanding the initial conditions in power series of  $\varepsilon$ , one gets:

$$\begin{aligned}\bar{x} &= \bar{x}_0 + \varepsilon \bar{x}_1 + \varepsilon^2 \bar{x}_2 + \dots \\ \bar{y} &= \bar{y}_0 + \varepsilon \bar{y}_1 + \varepsilon^2 \bar{y}_2 + \dots\end{aligned}\tag{28}$$

Inserting (27) and (28) in (24), equating same orders in  $\varepsilon$  and taking into account the periodicity condition (26), one can find the following explicit expressions for  $x_1(t), y_1(t), \bar{y}_0, \bar{y}_1$ :

$$\begin{aligned}y_1(t) &= y_1(t; \bar{y}, \bar{x}) = \int_0^t f_x(\bar{x}_0 + \bar{y}_0 s, s) ds \\ x_1(t) &= x_1(t; \bar{y}, \bar{x}) = \int_0^t y_1(s) ds \\ \bar{y}_0 &= \frac{p}{q} \\ \bar{y}_1 &= -\frac{1}{2\pi q} \int_0^{2\pi q} \int_0^t f_x(\bar{x}_0 + \bar{y}_0 s, s) ds dt.\end{aligned}$$

Furthermore,  $\bar{x}_0$  is determined as a solution of

$$\int_0^{2\pi q} f_x(\bar{x}_0 + \bar{y}_0 s, s) ds = 0,$$

while  $\bar{x}_1$  is given by

$$\bar{x}_1 = -\frac{1}{\int_0^{2\pi q} f_{xx}^0 dt} \cdot \left[ \bar{y}_1 \int_0^{2\pi q} t f_{xx}^0 dt + \int_0^{2\pi q} f_{xx}^0 x_1(t) dt \right],$$

where, for shortness, we have written  $f_{xx}^0 = f_{xx}(\bar{x}_0 + \bar{y}_0 t, t)$ .

### The Libration in Longitude of the Moon

The previous computation of the  $p : q$  periodic solution can be used to evaluate the libration in longitude of the Moon. More precisely, setting  $p = q = 1$ , one obtains

$$\begin{aligned} \bar{x}_0 &= 0 \\ \bar{y}_0 &= 1 \\ x_1(t) &= 0.232086t - 0.218318 \sin(t) \\ &\quad - 6.36124 \cdot 10^{-3} \sin(2t) - 3.21314 \cdot 10^{-4} \sin(3t) \\ &\quad - 1.89137 \cdot 10^{-5} \sin(4t) - 1.18628 \cdot 10^{-6} \sin(5t) \\ y_1(t) &= 0.232086 - 0.218318 \cos(t) - 0.0127225 \cos(2t) \\ &\quad - 9.63942 \cdot 10^{-4} \cos(3t) - 7.56548 \cdot 10^{-5} \cos(4t) \\ &\quad - 5.93138 \cdot 10^{-6} \cos(5t) \\ \bar{x}_1 &= 0 \\ \bar{y}_1 &= -0.232086, \end{aligned}$$

where we used  $e = 0.05494$ ,  $\varepsilon = 3.45 \cdot 10^{-4}$ . Therefore, the synchronous periodic solution, computed up to the first order in  $\varepsilon$ , is given by

$$\begin{aligned} x(t) &= \bar{x}_0 + \bar{y}_0 t + \varepsilon x_1(t) = t - 7.53196 \cdot 10^{-5} \sin(t) \\ &\quad - 2.19463 \cdot 10^{-6} \sin(2t) - 1.10853 \cdot 10^{-7} \sin(3t) \\ &\quad - 6.52523 \cdot 10^{-9} \sin(4t) - 4.09265 \cdot 10^{-10} \sin(5t) \\ y(t) &= \bar{y}_0 t + \varepsilon y_1(t) = 1 - 7.53196 \cdot 10^{-5} \cos(t) \\ &\quad - 4.38926 \cdot 10^{-6} \cos(2t) - 3.3256 \cdot 10^{-7} \cos(3t) \\ &\quad - 2.61009 \cdot 10^{-8} \cos(4t) - 2.04633 \cdot 10^{-9} \cos(5t). \end{aligned}$$

It turns out that the libration in longitude of the Moon, provided by the quantity  $x(t) - t$ , is of the order of  $7 \cdot 10^{-5}$  in agreement with the observational data.

### Future Directions

The last decade of the twentieth century has been greatly marked by astronomical discoveries, which changed the shape of the solar system as well as of the entourage of other stars. In particular, the detection of many small bodies beyond the orbit of Neptune has moved the edge of the solar system forward, and it has increased the number of its population. Hundreds of objects have been observed to move in a ring beyond Neptune, thus forming the so-called Kuiper's belt. Its components show a great variety of behaviors, like resonance clusterings, regular orbits, and scattered trajectories. Furthermore, far outside the solar system, the astronomical observations of extrasolar planetary systems have opened new scenarios with a great variety of dynamical behaviors. In these contexts, classical and resonant perturbation theories will deeply contribute to provide a fundamental insight of the dynamics and will play a prominent role in explaining the different configurations observed within the Kuiper's belt as well as within extrasolar planetary systems.

### Bibliography

- Andoyer H (1926) *Mécanique Céleste*. Gauthier-Villars, Paris
- Arnold VI (1963) Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat Nauk* 6 18(114):91–192
- Arnold VI (1978) *Mathematical methods of classical mechanics*. Springer, Berlin
- Arnold VI (ed) (1988) *Encyclopedia of mathematical sciences. Dynamical systems III*. Springer, Berlin
- Benettin G, Fasso F, Guzzo M (1998) Nekhoroshev-stability of  $L_4$  and  $L_5$  in the spatial restricted three-body problem. *Regul Chaotic Dyn* 3(3):56–71
- Boccaletti D, Pucacco G (2001) *Theory of orbits*. Springer, Berlin
- Brouwer D, Clemence G (1961) *Methods of celestial mechanics*. Academic Press, New York
- Celletti A (1990) Analysis of resonances in the spin-orbit problem. In: *Celestial mechanics: The synchronous resonance (Part I)*. *J Appl Math Phys (ZAMP)* 41:174–204
- Celletti A (1993) Construction of librational invariant tori in the spin-orbit problem. *J Appl Math Phys (ZAMP)* 45:61–80
- Celletti A, Chierchia L (1998) Construction of stable periodic orbits for the spin-orbit problem of Celestial

- Mechanics. Regul Chaotic Dyn (Editorial URSS) 3: 107–121
- Celletti A, Chierchia L (2006) KAM tori for N-body problems: a brief history. Celest Mech Dyn Astron 95(1): 117–139
- Celletti A, Chierchia L (2007) KAM stability and celestial mechanics. Mem Am Math Soc 187:878
- Celletti A, Giorgilli A (1991) On the stability of the Lagrangian points in the spatial restricted problem of three bodies. Celest Mech Dyn Astron 50:31–58
- Chebotaev AG (1967) Analytical and numerical methods of celestial mechanics. Elsevier, New York
- Chierchia L, Gallavotti G (1994) Drift and diffusion in phase space. Ann l'Inst H Poincaré 60:1–144
- Delaunay C (1867) Mémoire sur la théorie de la Lune. Mém l'Acad Sci 28:29. (1860)
- Deprit A (1967) Free rotation of a rigid body studied in the phase space. Am J Phys 35:424–428
- Eftymiopoulos C, Sandor Z (2005) Optimized Nekhoroshev stability estimates for the Trojan asteroids with a symplectic mapping model of co-orbital motion. MNRAS 364(6):253–271
- Féjóz J (2004) Démonstration du “théorème d’Arnold” sur la stabilité du système planétaire (d’après Michael Herman). Ergod Theory Dyn Syst 24:1–62
- Ferraz-Mello S (2007) Canonical perturbation theories. Springer, Berlin
- Gabern F, Jorba A, Locatelli U (2005) On the construction of the Kolmogorov normal form for the Trojan asteroids. Nonlinearity 18:1705–1734
- Giorgilli A, Skokos C (1997) On the stability of the trojan asteroids. Astron Astrophys 317:254–261
- Giorgilli A, Delshams A, Fontich E, Galgani L, Simó C (1989) Effective stability for a Hamiltonian system near an elliptic equilibrium point, with an application to the restricted three-body problem. J Differ Equ 77:167–198
- Hagihara Y (1970) Celestial mechanics. MIT Press, Cambridge
- Hénon M (1966) Explorations numériques du problème restreint IV: Masses égales, orbites non périodique. Bull Astron 3(1, fasc 2):49–66
- Kolmogorov AN (1954) On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian. Dokl Akad Nauk SSR 98: 527–530
- Laskar J, Robutel P (1995) Stability of the planetary three-body problem I expansion of the planetary Hamiltonian. Celest Mech Dyn Astron 62(3):193–217
- Lhotka C, Efthymiopoulos C, Dvorak R (2008) Nekhoroshev stability at  $L_4$  or  $L_5$  in the elliptic restricted three body problem-application to Trojan asteroids. MNRAS 384:1165–1177
- Locatelli U, Giorgilli A (2000) Invariant tori in the secular motions of the three-body planetary systems. Celest Mech Dyn Astron 78:47–74
- Locatelli U, Giorgilli A (2005) Construction of the Kolmogorov’s normal form for a planetary system. Regul Chaotic Dyn 10:153–171
- Locatelli U, Giorgilli A (2007) Invariant tori in the Sun-Jupiter-Saturn system. Discret Contin Dynam Syst-Ser B 7:377–398
- Meyer KR, Hall GR (1991) Introduction to Hamiltonian dynamical systems and the N-body problem. Springer, Berlin
- Moser J (1962) On invariant curves of area-preserving mappings of an annulus. Nach Akad Wiss Göttingen Math Phys Kl II 1:1
- Poincaré H (1892) Les Méthodes Nouvelles de la Mécanique Céleste. Gauthier-Villars, Paris
- Robutel P (1995) Stability of the planetary three-body problem II KAM theory and existence of quasi-periodic motions. Celest Mech Dyn Astron 62(3): 219–261
- Robutel P, Gabern F (2006) The resonant structure of Jupiter’s Trojan asteroids – I long-term stability and diffusion. MNRAS 372(4):1463–1482
- Sanders JA, Verhulst F (1985) Averaging methods in nonlinear dynamical systems. Springer, Berlin
- Siegel CL, Moser JK (1971) Lectures on celestial mechanics. Springer, Heidelberg
- Szebehely V (1967) Theory of orbits. Academic Press, New York

# n-Body Problem and Choreographies

Susanna Terracini  
Dipartimento di Matematica “Giuseppe Peano”,  
Universita di Torino, Turin, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Simple Choreographies and Relative Equilibria  
Symmetry Groups and Equivariant Orbits  
The 3-Body Problem  
Minimizing Properties of Simple Choreographies  
Generalized Orbits and Singularities  
Asymptotic Estimates at Collisions  
Absence of Collision for Locally Minimal Paths  
Future Directions  
Bibliography

## Glossary

**Central configurations** Are the critical points of the potential constrained on the unitary moment of inertia ellipsoid. Central configurations are associated to particular solutions to the  $n$ -body problem: the *relative equilibrium* and the *homographic* motions defined in Definition 1.

**Choreographical solution** A choreographical solution of the  $n$ -body problem is a solution such that the particles move on the same curve, exchanging their positions after a fixed time. This property can be regarded as a symmetry of the trajectory. This notion finds a natural generalization in that of  $G$ -equivariant trajectory defined in Definition 2 for a given group of symmetries- $G$ . The  $G$ -equivariant minimization technique consists in seeking action minimizing trajectories among all  $G$ -equivariant paths.

**Collision and singularities** When a trajectory can not be extended beyond a certain time  $b$  we say that a *singularity* occurs. Singularities can be *collisions* if the solution admits a limit configuration as  $t \rightarrow b$ . In such a case we term  $b$  a *collision instant*.

**$n$ -Body problem** The  $n$ -body problem is the system of differential eqs. (1) associated with suitable initial or boundary value data. A *solution* or *trajectory* is a doubly differentiable path  $q(t) = (q_1(t), \dots, q_n(t))$  satisfying (1) for all  $t$ . The weaker notion of *generalized solution* is defined in Definition 5 applies to trajectories found by variational methods.

**Variational approach** The variational approach to the  $n$ -body problem consists in looking at trajectories as critical points of the *action functional* defined in (4). Such critical points can be (local) *minimizers*, or *constrained minimizers* or *mountain pass*, or other type.

## Definition of the Subject

The motion of  $n$ -point particles of positions  $x_i(t) \in \mathbb{R}^3$  and masses  $m_i > 0$ , interacting in accordance with Newton’s law of gravitation, satisfies the system of differential equations:

$$-m_i \ddot{x}_i(t) = G \sum_{j \neq i, j=1}^n m_i m_j \frac{x_i - x_j}{|x_i - x_j|^3}, \quad (1)$$

$$i = 1, \dots, n, \quad t \in \mathbb{R}.$$

The  $n$ -body problem consists in solving Eq. (1) associated with initial or boundary conditions.

A simple choreography is a periodic solution to the  $n$ -body problem Eq. (1) where the bodies lie on the same curve and exchange their mutual positions after a fixed time, namely, there exists a function  $x : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$x_i(t) = x(t + (i - 1)\tau), \quad (2)$$

$$i = 1, \dots, n, \quad t \in \mathbb{R},$$

where  $\tau = 2\pi/n$ .



## Introduction

The two-body problem can be reduced, by the conservation of the linear momentum, to the *one center* Kepler problem and can be completely solved either by exploiting the conservation laws (angular momentum, energy and the Lenz vector), or by performing the Levi–Civita change of coordinates reducing the problem to that of an harmonic oscillator (Levi Civita 1918).

The three-body problem is much more complicated than the two-body and can not be solved in a simple way. A major study of the Earth–Moon–Sun system was undertaken by Delaunay in his *La Théorie du mouvement de la lune*. In the restricted three-body problem, the mass of one of the bodies is negligible; the circular restricted three-body problem is the special case in which two of the bodies are in circular orbits and was worked on extensively by many famous mathematicians and physicists, notably Lagrange in the eighteenth century, Poincaré at the end of the nineteenth century and Moser in the twentieth century. Poincaré’s work on the restricted three-body problem was the foundation of deterministic chaos theory.

A very basic – still fundamental – question concerns the real number of degrees of freedom of the  $n$ -body problem. As the motion of each point particle is represented by 3-dimensional vectors, the  $n$ -body problem has  $3n$  degrees of freedom and hence it is  $6n$ -dimensional. First integrals are: the center of mass, the linear momentum, the angular momentum, and the energy. Hence there are 10 independent algebraic integrals. This allows the reduction of variables to  $6n - 10$ . It was proved in 1887 by Bruns that these are the only linearly independent integrals of the  $n$ -body problem, which are algebraic with respect to phase and time variables. This theorem was later generalized by Poincaré.

A second very natural problem is whether there exists a power series expressing – if not every – at least a large and relevant class of trajectories for the  $n$ -body problem. In 1912, after pioneering works of Mittag–Leffler and Levi–Civita, Sundman proved the existence of a series solution in powers of  $t^{1/3}$  for the 3-body problem. This

series is convergent for all real  $t$ , except for those initial data which correspond to *vanishing angular momentum*. These initial data have Lebesgue measure zero and therefore are not generic. An important issue in proving this result is the fact that the radius of convergence for this series is determined by the distance to the nearest singularity. Since then, the study of singularities became the main point of interest in the study of the  $n$ -body problem. Sundman’s result was later generalised to the case of  $n > 3$  bodies by Q Wang in (Wang 1991). However, the rate and domain of convergence of this series are so limited to make it hardly applicable to practical and theoretical purposes.

Finally, the  $n$ -body problem can be faced from the point of view of the theory of perturbations and represents both its starting point and its most relevant application. Delaunay and Poincaré described the spatial three-body problem as a four-dimensional Hamiltonian system and already encountered some trajectories featuring a chaotic behavior. When one of the bodies is much heavier than the other two (a system of one “star” and two “planets”), one can neglect the interaction between the small planets; hence the system can be seen as a perturbation of two decoupled two-body problems whose motions are known to be all periodic from Kepler’s laws. In the planetary three-body problem, hence, two harmonic oscillators interact nonlinearly through the perturbation. Resonances between the two oscillators can be held responsible of high sensitivity with respect to initial data and other chaotic features. A natural question regards the coexistence of the irregular trajectories with regular (periodic or quasi-periodic) ones. A modern approach to the problem of stability of solutions to nearly integrable systems goes through the application of Kolmogorov–Arnold–Moser (KAM) Theorem (Arnold 1963; Jefferys and Moser 1966), whose main object indeed is indeed the persistence, under perturbations, of invariant tori.

The  $n$ -body problem is paradigmatic of any complex system of many interacting objects, and it can neither be solved nor it can be simplified in an efficient way. A possible starting point for its analysis is to seek selected trajectories whose

motion is particularly simple, in the sense that it repeats after a fixed period: the periodic solutions. Following Poincaré in his *Méthodes Nouvelles de la Mécanique Céleste*, tome I, 1892,

... D'ailleurs, ce qui nous rends ces solutions périodiques si précieuses, c'est qu'elles sont, pour ainsi dire, la seule brèche par où nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable.

Indeed, just before, Poincaré conjectured that periodic trajectories are dense in the phase space:

... Voici un fait que je n'ai pas pu démontrer rigoureusement, mais qui me paraît pourtant très vraisemblable. Étant données des équations de la forme définie dans le n.13 (Formula N. 13 quoted by Poincaré is Hamilton equation and covers our class of Dynamical Systems Eq. (3)) et une solution quelconque de ces équations, on peut toujours trouver une solution périodique (dont la période peut, il est vrai, être très longue), telle que la différence entre les deux solutions soit aussi petite qu'on le veut, pendant un temps aussi long qu'on le veut.

### Singular Hamiltonian Systems

From an abstract point of view, the  $n$ -body problem is a Hamiltonian System of the form

$$m_i \ddot{x}_i = \frac{\partial U}{\partial x_i}(t, x), \quad i = 1, \dots, n, \quad (3)$$

where the forces  $\frac{\partial U}{\partial x_i}$  are undefined on a singular set  $\Delta$ , the set of collisions between two or more particles in the  $n$ -body problem. Such singularities play a fundamental role in the phase portrait (see, e.g. (Diacu 2002)) and strongly influence the global orbit structure, as they can be held responsible, among others, of the presence of chaotic motions (see, e.g. (Devaney 1981)) and of *motions becoming unbounded in a finite time* (Mather and McGehee 1974; Xia 1992).

Two are the major steps in the analysis of the impact of the singularities in the  $n$ -body problem: the first consists in performing the asymptotic analysis along a single collision (total or partial) trajectory and goes back, in the classical case, to the works by (Sundman 1913), Wintner (1941) and, in more recent years by Sperling, Pollard, Saari and other authors (see for instance (Diacu 1992; ElBialy 1990; Pollard and Saari 1968, 1970; Saari 1972; Sperling 1970)). The second

step consists in blowing-up the singularity by a suitable change of coordinates introduced by McGehee in (McGehee 1974) and replacing it by an invariant boundary – the collision manifold – where the flow can be extended in a smooth manner. It turns out that, in many interesting applications, the flow on the collision manifold has a simple structure: it is a gradient-like, Morse–Smale flow featuring a few stationary points and heteroclinic connections (see, for instance, the surveys (Devaney 1981; Moeckel 1987)). The analysis of the extended flow allows us to obtain a full picture of the behavior of solutions near the singularity, despite the flow fails to be fully regularizable (except for binary collisions).

### Simple Choreographies and Relative Equilibria

A possible starting point for the study of the  $n$ -body problem is to find selected trajectories which are particularly simple, when regarded from some point of view. Examples of such particular solutions are the collinear periodic orbits (found in 1767 by Euler), in which three bodies of any masses move such that they oscillate along a rotation line, and the Lagrange triangular solutions, where the bodies lie at the vertices of a rotating equilateral triangle that shrinks and expands periodically, discovered in 1772. Both these trajectories are stationary in a rotating frame. A second remarkable class of trajectories – the choreographies – can be found when the masses are all equal, by exploiting the fact that particles can be interchanged without changing the structure of the system.

Among all periodic solutions of the planar 3-body problem, the relative equilibrium motions – the equilateral Lagrange and the collinear Euler–Moulton solutions – are definitely the simplest and most known. In general such simple periodic motions exist for any number of bodies.

**Definition 1** A *relative equilibrium* trajectory is a solution of (1) whose configuration remains constant in a rotating frame. A *homographic* trajectory is a solution of (1) whose configuration remains constant up to homotheties.

The normalized configurations of such trajectories are named *central* and are the critical points of the potential

$$U(x) = \sum_{i < j} \frac{m_i m_j}{|x_i - x_j|},$$

constrained to the ellipsoid of unitary momentum of inertia:

$$I(x) = \sum_i m_i |x_i|^2 = 1.$$

Relative equilibria feature an evident symmetry ( $SO(2)$  and  $O(2)$  respectively), that is, they are equivariant with respect to the symmetry group of dimension 1 acting as  $SO(2)$  (resp.  $O(2)$ ) on the time circle and on the plane, and trivially on the set of indexes  $\{1, 2, 3\}$ . In fact, they are minimizers of the Lagrangian action functional in the space of all loops having their same symmetry group. Therefore, generally speaking, *G-equivariant minimizers for the action functional (given a symmetry group G) can be thought as the natural generalization of relative equilibrium motions.*

This perspective has known a wide popularity in the recent literature and has produced a new boost in the study of periodic trajectories in the  $n$ -body problem; the recent proof of the existence of the Chenciner–Montgomery eight-shaped orbit is emblematic of this renewed interest (see (Arioli et al. 2006; Chen 2001a, 2003a; Chenciner 2002b; Chenciner and Montgomery 2000; Ferrario and Terracini 2004) and the major part of our bibliographical references). In all these papers, periodic and quasi-periodic solutions of the  $n$ -body problem can be found as critical points of the Lagrangian action functional restricted to suitable spaces of symmetric paths.

### Basic Definitions and Notations

Let us consider  $n$  point particles with masses  $m_1, m_2, \dots, m_n$  and positions  $x_1, x_2, \dots, x_n \in \mathbb{R}^d$ , with  $d \geq 2$ . We denote by  $X$  the space of configurations with center of mass in 0, and by  $\hat{X} = X \setminus \Delta$  the set of collision-free configurations (collision means  $x_i = x_j$  for some  $i \neq j$ ). On the configuration space we define the homogeneous (Newton) potential of degree  $-\alpha < 0$ :

$$U(x) = \sum_{i < j} U_{ij}(|x_i - x_j|), \quad U_{ij}(|x_i - x_j|) \simeq \frac{m_i m_j}{|x_i - x_j|^\alpha}.$$

In many cases one can simply require the  $U_{i,j}$  to be asymptotically homogeneous at the singularity. Furthermore, a major part of our analysis can be extended to logarithmic potentials. Relative equilibria correspond to those configurations (termed central) which are critical for the restriction of the potential  $U$  to the ellipsoid  $I = \lambda$  where  $I$  denotes the momentum of inertia:

$$I(x) = \sum_i m_i |x_i|^2.$$

On collisions the potential  $U = +\infty$ . We are interested in (relative) periodic (such that  $\forall t : x(t+T) = x(t)$ ) solutions to the system of differential equations:

$$m_i \ddot{x}_i = \frac{\partial U}{\partial x_i}.$$

We associate with the equation the Lagrangian integrand

$$\begin{aligned} L(x, \dot{x}) &= L = K + U \\ &= \sum_i \frac{1}{2} m_i |\dot{x}_i|^2 + \sum_{i < j} U_{ij}(|x_i - x_j|) \end{aligned}$$

and the action functional:

$$\mathcal{A}(x) = \int_0^T L(x(t), \dot{x}(t)) dt. \quad (4)$$

Sometimes it will be preferable to consider the problem in a frame rotating uniformly about the vertical axis, with an angular speed  $\omega$ ; the corresponding action  $\mathcal{A}^\omega$  then contains a gyroscopic term, associated with Coriolis force.

We shall seek periodic solutions as critical points of the action functional on the Sobolev space of  $T$ -periodic trajectories:  $A = H^1(\mathbb{T}, X)$  or, to be more precise, of the action constrained on suitable linear subspaces  $A_0 \subset A$ .

Two are the major difficulties to be faced in following the variational approach; the first is due

to the *lack of coercivity* (or of Palais–Smale) due to the vanishing at infinity of the force fields: indeed sequences of almost-critical points (such as minimizing sequences) may very well diverge. Furthermore, as the potential  $U$  is singular on collisions, minimizers or other critical points can a priori be *collision trajectories*. Compactness can be successfully recovered by the symmetry of the problem, as we are going to explain here below. We shall expose in the last part of this article some of the strategies which has been developed in order to overcome the problem of collisions.

## Symmetry Groups and Equivariant Orbits

We can generalize the concept of relative equilibria as follows: we impose the permutation of the positions of the particles, after a given time, up to isometries of the space. This gives rise to a class of symmetric trajectories (the generalized choreographies). It is worthwhile noticing that these trajectories arise as a common feature of any system of interacting objects, regardless whether they come from models in Celestial mechanics or not. It applies to atoms, or molecules, galaxies or whatever system, provided the objects can be freely interchanged. Let us start by introducing some basics concepts and definitions from (Ferrario and Terracini 2004). Let  $G$  be a finite group endowed with:

- an orthogonal representation of dimension  $2$   $\tau : G \rightarrow O(2)$  (on cyclic time  $\mathbb{T} \cong S^1$ ),
- an orthogonal representation (on the euclidean space  $\mathbb{R}^d$ )  $\rho : G \rightarrow O(d)$ ,
- and an homomorphism on the symmetric group on  $n$  elements ( $\mathbf{n} = \{1, 2, \dots, n\}$ )  $\sigma : G \rightarrow \Sigma_n$ .

Then  $G$  acts on time (translation and reversal)  $\mathbb{T}$  via  $\tau$  and it acts on the configuration space  $\mathcal{X}$  via  $\rho$  and  $\sigma$  in the following way

$$\forall i = 1 \dots n : (gx)_i = \rho(g)x_{\sigma(g)^{-1}(i)}.$$

As a consequence we have an action on the space of trajectories:

**Definition 2** A continuous function  $x(t) = (x_1(t), \dots, x_n(t))$  is *G-equivariant* if

$$\forall g \in G : x(gt) = (gx)(t).$$

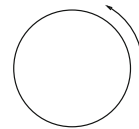
The linear subspace  $A_0 = A^G \subset A$  denotes the set of periodic trajectories in? which are equivariant with respect to the  $G$ -action:

The Palais Principle of symmetric criticality (Palais 1979) ensures that critical points of an invariant functional restricted to the space of equivariant trajectories are indeed free critical points. We stress that by the particular form of the interaction potentials, in our setting invariance is simply implied by equality of those masses which are interchanged by the action of  $G$  on the set of the indices. When the action of  $s$  is transitive, all the masses must be equal and the associated  $G$ -equivariant trajectories give rise to *generalized choreographies*.

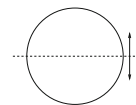
## Cyclic and Dihedral Actions

Consider the normal subgroup  $\ker \tau \triangleleft G$  and the quotient  $\overline{G} = G/\ker \tau$ . Since  $\overline{G}$  acts effectively on  $\mathbb{T}$ , it is either a *cyclic* group or a *dihedral* group.

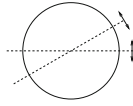
- If the group  $\overline{G}$  acts trivially on the orientation of  $\mathbb{T}$ , then  $\overline{G}$  is cyclic and we say that the action of  $G$  on  $A$  is of *cyclic type*.



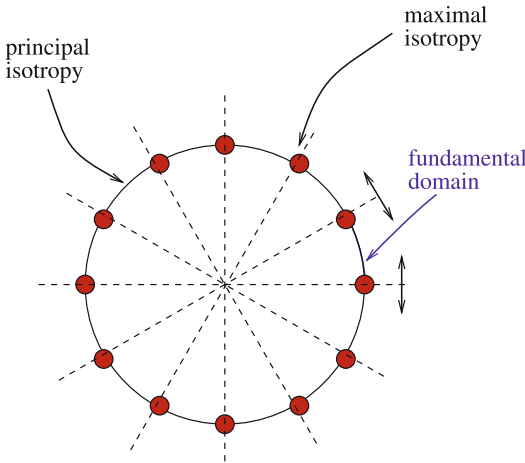
- If the group  $\overline{G}$  consists of a single reflection on  $\mathbb{T}$ , then we say that action of  $G$  on  $A$  is of *brake type*.



- Otherwise, we say that the action of  $G$  on  $A$  is of *dihedral type*.



With this first classification, we can easily associate to our minimization problem some proper boundary conditions.



**Proposition 1** Let  $\mathbb{I}$  be the *fundamental domain* (for a dihedral type), or any interval having as length the minimal angle of time-rotation in  $\overline{G}$  (for a cyclic type). Then the  $G$ -equivariant minimization problem is equivalent to problem of minimizing the action over all paths  $x : \mathbb{I} \rightarrow \mathcal{X}^{\ker \tau}$  subject to the boundary conditions  $x(0) \in \mathcal{X}^{H_0}$  and  $x(1) \in \mathcal{X}^{H_1}$ , where  $H_0$  and  $H_1$  are the maximal isotropy subgroups of the boundary of  $\mathbb{I}$ .

### The Variational Approach

Let us consider the  $\mathcal{A}$  restricted to the space of symmetric loops  $\mathcal{A}^G$ . We recall that the action functional is said to be coercive if  $\lim_{|x| \rightarrow +\infty} \mathcal{A}(x) = +\infty$ . Coercivity implies the validity of the direct method of Calculus of Variations and, consequently, the existence of a minimizer.

**Proposition 2** The action functional  $\mathcal{A}$  is coercive in  $\mathcal{A}^G$  if and only if  $\chi^G = 0$ . Consequently, if  $\chi^G = 0$  then a minimizer of  $\mathcal{A}^G$  in  $\mathcal{A}^G$  exists.

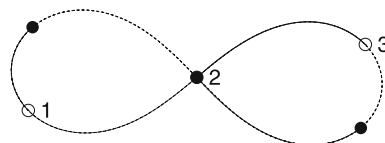
Given  $\rho$  and  $\sigma$ , we can compute  $\dim \chi^G$

$$\dim \chi^G = \frac{1}{|G|} \sum_{g \in G} \text{Tr}(\rho(g)) \# \text{Fix}(\sigma(g)) - d.$$

In the frame rotating with constant angular speed  $\omega$  the action  $\mathcal{A}^\omega$  is generally coercive, except for a (possible) discrete set values of  $\omega$ .

Of course, beside seeking minimizers, one can look for other type of critical points, such as mountain pass or others. As the potential  $U$  is singular on collisions, minimizers (or other critical points) can a priori be *collision trajectories*. Many strategies were proposed in the literature in order to overcome these obstacles. The development of a suitable Critical Point theory taking into account of the contribution of fake periodic solutions (the critical points at infinity) was proposed by some authors (Bahri and Rabinowitz 1991; Majer and Terracini 1995a; Riahi 1999) and returned a good estimate of the number of periodic trajectories satisfying an appropriate bound on the length, with the main disadvantage of requiring a strong order of infinity at the collisions (the *strong force condition*). Another strategy to recover coercivity of action functional consists in imposing a symmetry constraint on the loop space. Surprisingly enough, once coercivity is recovered, also the problem of collisions becomes much less dramatic. This fact was remarked for the first time in Degiovanni et al. (1987) and widely exploited in the literature, also thanks to the neat idea, due to C. Marchal, of averaging over all possible variations (generalized and exposed in section “Generalized Orbits and Singularities”) to avoid the occurrence of collisions for extremals of the action (Marchal idea was first exposed in Chenciner (2002b)). This argument can be used in most of the known cases to prove that absence of collisions for minimizing trajectories and will be outlined in the last section.

### The Eight Shaped Three-Body Solution



In their paper (Chenciner and Montgomery 2000), Chenciner and Montgomery exploited a variational argument in order to prove the existence of a periodic trajectory for the three body problem, where the three particles move one a single eight-shaped curve, interchanging their positions after a fixed time. C. Moore (1993) was the first to find numerically the eight, lead by topological reasons which turned out to be insufficient to insure its existence. One of the the simplest symmetries giving rise to the eight shaped trajectory is the following. Denote

$$x(t) = (x_1(t), x_2(t), x_3(t)) \in \mathbb{R}^6.$$

Let  $G = D_6$  be the dihedral group generated by two following reflections: the first

$$\begin{aligned} x_1(-t) &= -x_1(t), & x_2(-t) &= -x_3(t), \\ x_3(-t) &= -x_2(t), \end{aligned}$$

with  $g_1 = (\tau_1, \rho_1, \sigma_1)$  is  $\tau_1(t) = -t$ ,  $\rho_1(x) = -x$  and  $\sigma_1(1, 2, 3) = (1, 3, 2)$ . And the second

$$\begin{aligned} x_1(1-t) &= -x_2(t), & x_2(1-t) &= -x_1(t), \\ x_3(1-t) &= -x_3(t), \end{aligned}$$

with  $g_2 = (\tau_2, \rho_2, \sigma_2)$  is  $\tau_2(t) = 1-t$ ,  $\rho_2(x) = -x$  and  $\sigma_2(1, 2, 3) = (2, 1, 3)$ .

There are three possible groups yielding an eight-shaped trajectory. First, we consider the group of cyclic action type  $C_6$  (the *cyclic eight* having order 6), which acts cyclically on  $\mathbb{T}$  (i.e. by a rotation of angle  $\pi/3$ ), by a reflection in the plane  $E$ , and by the cyclic permutation  $(1, 2, 3)$  in the index set.

The second group, which we denote by  $D_{12}$ , is the group of order 12 obtained by extending  $C_6$  with the element  $h$  defined as follows:  $\tau(h)$  is a reflection in  $\mathbb{T}$ ,  $\rho(h)$  is the antipodal map in  $E$  (thus, the rotation of angle  $\pi$ ), and  $\sigma(h)$  is the permutation  $(1, 2)$ . This is the symmetry group used by Chenciner and Montgomery.

The third group is the subgroup of  $D_{12}$  generated by  $h$  and the subgroup  $C_3$  of order 3 of  $C_6 \subset D_{12}$ . We denote this group  $D_6$  (since it is a dihedral group of order 6). The symmetry

groups  $D_{12}$  and  $D_6$  are of dihedral type. The choreography group  $C_3$  is a subgroup of all the three groups, thus the action is coercive on  $G$ -equivariant loops.

### The Rotating Circle Property (RCP)

As a second step of the variational approach to the  $n$ -body problem, one has to prove that the output of the minimization (or some other variational method) procedure is free of collisions. This point involves a deep analysis of the structure of the possible singularities and will be outlined in the last part of this article. After performing the analysis of collisions, we find that *if the action of  $G$  on  $\mathbb{T}$  and  $X$  fulfills some conditions (computable) then (local) minimizers of the action functional  $\mathcal{A}$  in  $A^G \subset A$  do not have collisions.*

For a group  $H$  acting orthogonally on  $\mathbb{R}^d$ , a circle  $\mathbb{S} \subset \mathbb{R}^d$  (with center in 0) is termed *rotating under  $H$*  if  $\mathbb{S}$  is *invariant* under  $H$  (that is, for every  $g \in H$   $g\mathbb{S} = \mathbb{S}$ ) and for every  $g \in H$  the restriction  $g|_{\mathbb{S}} : \mathbb{S} \rightarrow \mathbb{S}$  is a *rotation* (the identity is meant as a  $\sim$ -rotation of angle 0).

Let  $i \in \mathbf{n}$  be an index and  $H \subset G$  a subgroup. A circle  $\mathbb{S} \subset \mathbb{R}^d = V$  (with center in 0) is termed *rotating for  $i$  under  $H$*  if  $\mathbb{S}$  is *rotating under  $H$*  and

$$\mathbb{S} \subset V^{H_i} \subset V = \mathbb{R}^d,$$

where  $H_i \subset H$  denotes the *isotropy subgroup* of the index  $i$  in  $H$  relative to the action of  $H$  on the index set  $\mathbf{n}$  induced by restriction (that is, the isotropy  $H_i = \{g \in H \mid gi = i\}$ ).

**Definition 3** A group  $G$  acts with the *rotating circle property* if for every  $\mathbb{T}$ -isotropy subgroup  $G_i \subset G$  and for at least  $n-1$  indexes  $i \in \mathbf{n}$  there exists in  $\mathbb{R}^d$  a rotating circle  $\mathbb{S}$  under  $G_i$  for  $i$ .

In most of the known examples the property is fulfilled. In Ferrario and Terracini (2004) the following results were proved.

**Theorem 1** Consider a finite group  $K$  acting on? with the rotating circle property. Then a minimizer of the  $K$ -equivariant fixed-ends (Bolza) problem is free of collisions.



**Corollary 1** For every  $\alpha > 0$ , minimizers of the fixed-ends (Bolza) problem are free of interior collisions.

**Corollary 2** If the action of  $G$  on  $A$  is of cyclic type and  $\ker\tau$  has the rotating circle property then any local minimizer of  $\mathcal{A}^G$  in  $A^G$  is collisionless.

**Corollary 3** If the action of  $G$  on  $A$  is of cyclic type and  $\ker\tau = 1$  is trivial then any local minimizer of  $\mathcal{A}^G$  in  $A^G$  is collisionless.

**Theorem 2** Consider a finite group  $G$  acting on  $A$  so that every maximal  $\mathbb{T}$ -isotropy subgroup of  $G$  either has the rotating circle property or acts trivially on the index set  $\mathbf{n}$ . Then any local minimizer of  $\mathcal{A}^G$  yields a collision-free periodic solution of the Newton equations for the  $n$ -body problem in  $\mathbb{R}^d$ .

### Examples

In this section, for the sake of illustrating the power and the limitations of the approach through the  $G$ -equivariant minimization, we include a few examples fitting in our theoretical framework and we give some hints of which symmetry groups satisfy or not assumptions of Theorem 2. Well-known examples are the celebrated Chenciner–Montgomery “eight” (Chenciner and Montgomery 2000), Chenciner–Venturelli “Hip–Hop” solutions (Chenciner and Venturelli 2000), Chenciner “generalized Hip–Hops” solutions (Chenciner 2002c), Chen’s orbit (Chen 2001a, 2003a) and Terracini–Venturelli generalized Hip–Hops (Terracini and Venturelli 2007). One word about the pictures of planar orbits: the configurations at the boundary points of the fundamental domain  $\mathbb{I}$  are denoted with an empty circle (starting point  $x_i(0)$ ) and a black disc (ending point  $x_i(t)$ , with  $t$  appropriate), with a label on the starting point describing the index of the particle. The trajectories of the particles with the times in  $\mathbb{I}$  are painted as thicker lines (thus it is possible to recover the direction of the movement from  $x_i(0)$  to  $x_i(t)$ ). Unfortunately this feature was not possible with the three-dimensional images.

Also, in all the following examples but 4 and 5 existence of the orbits follows directly from the

results of Theorem 2. The existence of the orbits described in Examples 4 and 5, which goes beyond the scope of this article, has been recently proved by Chen in (Chen 2003a). Thousands of other suitable actions and the corresponding orbits have been found by a special-purpose computer program based on GAP (The GAP Group 2002).

**Example 1 (Choreographies)** Consider the cyclic group  $G = \mathbb{Z}_n$  of order  $n$  acting trivially on  $V$ , with a cyclic permutation of order  $n$  on the index set  $\mathbf{n} = \{1, \dots, n\}$  and with a rotation of angle  $2\pi/n$  on the time circle  $\mathbb{T}$ . Since  $\chi^G = 0$ , by Proposition 2 the action functional  $\mathcal{A}^G$  is coercive. Moreover, since the action of  $G$  on  $\mathbb{T}$  is of cyclic type and  $\ker\tau = 1$ , by Corollary 2 the minimum exists and it has no collisions. For several numerical results and a description of choreographies we refer the reader to (Chenciner et al. 2001). An insight of the variational properties of choreographies is provided in section “Minimizing Properties of Simple Choreographies”.

**Example 2** Let  $n$  be odd. Consider the dihedral group  $G = D_{2n}$  of order  $2n$ , with the presentation

$$G = \langle g_1, g_2 \mid g_1^2 = g_2^n = (g_1 g_2)^2 = 1 \rangle.$$

Let  $\tau$  be the homomorphism defined by

$$\begin{aligned} \tau(g_1) &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{ and } \tau(g_2) \\ &= \begin{bmatrix} \cos \frac{2\pi}{n} & -\sin \frac{2\pi}{n} \\ \sin \frac{2\pi}{n} & \cos \frac{2\pi}{n} \end{bmatrix}. \end{aligned}$$

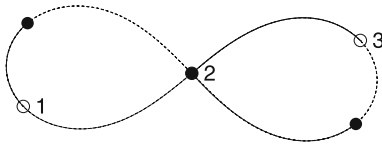
Furthermore, let the homomorphism? be defined by

$$\rho(g_1) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \text{ and } \rho(g_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

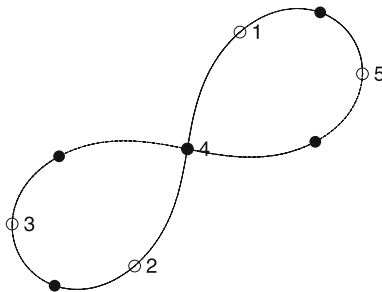
Finally, let  $G$  act on  $\mathbf{n}$  by the homomorphism  $s$  defined as  $\sigma(g_1) = (1, n-1)(2, n-2) \dots ((n-1)/2, (n+1)/2)$ ,  $\sigma(g_2) = (1, 2, \dots, n)$ , where

$(i_1, i_2, \dots, i_k)$  means the standard cycle-decomposition notation for permutation groups. By the action of  $g_2$  it is easy to show that all the loops in  $\Lambda^G$  are choreographies, and thus that, since  $\chi^G = 0$ , the action functional is coercive. The maximal  $\mathbb{T}$ -isotropy subgroups are the subgroups of order 2 generated by the elements  $g_1 g_2^i$  with  $i = 0, \dots, n-1$ . Since they are all conjugated, it is enough to show that one of them acts with the rotating circle property. Thus consider  $H = \langle g_1 \rangle \subset G$ . For every index  $i \in \{1, 2, \dots, n-1\}$  the isotropy  $H_i \subset H$  relative to the action of  $H$  on  $\mathbf{n}$  is trivial, and  $g_1$  acts by rotation on  $V = \mathbb{R}^2$ . Therefore for every  $i \in \{1, 2, \dots, n-1\}$  it is possible to choose a circle rotating under  $H$  for  $i$ , since, being  $H_i$  trivial,  $V^{H_i} = V$ . The resulting orbits are not homographic (since all the particles pass through the origin 0 at some time of the trajectory and the configurations are centered). For  $n = 3$  this is the *eight with less symmetry* of (Chenciner 2002b). Possible trajectories are shown in Figs. 1 and 2.

**Example 3** As in the previous example, let  $n \geq 3$  be an odd integer. Let  $G = C_{2n} \cong \mathbb{Z}_2 + \mathbb{Z}_n$  be the cyclic group of order  $2n$ , presented as  $G = \langle g_1, g_2 | g_1^2 = g_2^n = g_1 g_2 g_1^{-1} g_2^{-1} = 1 \rangle$ . The action



**n-Body Problem and Choreographies, Fig. 1** The ( $D_6$ -symmetric) eight for  $n = 3$



**n-Body Problem and Choreographies, Fig. 2** The ( $D_{10}$ -symmetric) eight with  $n = 5$

of  $G$  on  $\mathbb{T}$  is given by  $\tau(g_1 g_2) = \theta_{2n}$ , where  $\theta_{2n}$  denotes the rotation of angle  $\pi/n$  (hence the action will be of cyclic type). Now,  $G$  can act on the plane  $V = \mathbb{R}^2$  by the homomorphism defined by

$$\rho(g_1) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \rho(g_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Finally, the action of  $G$  on  $\mathbf{n} = \{1, 2, \dots, n\}$  is given by the homomorphism  $\sigma : G \rightarrow \Sigma_n$  defined by  $\sigma(g_1) = ()$ ,  $\sigma(g_2) = (1, 2, \dots, n)$ . The cyclic subgroup  $H_2 = \langle g_2 \rangle \subset G$  gives the symmetry constraints of the choreographies, hence loops in  $\Lambda^G$  are choreographies and the functional is coercive. Furthermore, since the action is of cyclic type, by Corollary 3 the minimum of the action functional is collisionless. It is possible that such minima coincide with the minima of the previous example: this would imply that the symmetry group of the minimum contains the two groups above.

**Example 4** Consider four particles with equal masses and an odd integer  $q \geq 3$ . Let be the direct product of the dihedral group of order  $4q$  with the group  $C_2$  of order 2. Let  $D_{4q}$  be presented by  $D_{4q} = \langle g_1, g_2 | g_1^2 = g_2^{2q} = (g_1 g_2)^2 = 1 \rangle$ , and let  $c \in C_2$  be the non-trivial element of  $C_2$ . Now define the homomorphisms  $\rho$ ,  $\tau$  and  $\sigma$  as follows:

$$\rho(g_1) = \tau(g_1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\rho(g_2) = \tau(g_2) = \begin{bmatrix} \cos \frac{2\pi}{2q} & -\sin \frac{2\pi}{2q} \\ \sin \frac{2\pi}{2q} & \cos \frac{2\pi}{2q} \end{bmatrix},$$

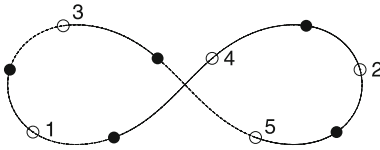
$$\rho(c) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \tau(c) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\sigma(g_1) = (1, 2)(3, 4), \quad \sigma(g_2) = (1, 3)(2, 4), \\ \sigma(c) = (1, 2)(3, 4).$$

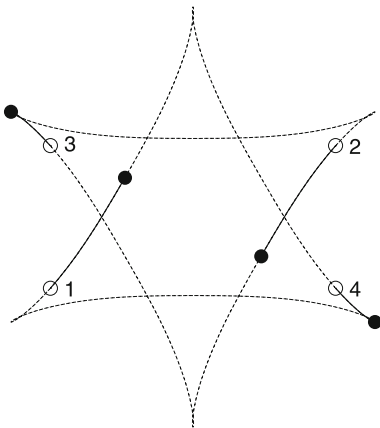
It is not difficult to show that  $\chi^G = 0$ , and thus the action is coercive. Moreover,  $\ker \tau = C_2$ ,

which acts on  $\mathbb{R}^2$  with the rotation of order 2, hence  $\ker \tau$  acts with the rotating circle property. Thus, by Proposition 2 and Theorem 1 the minimizer exists and does not have interior collisions. To exclude boundary collisions we cannot invoke Theorem 2, since the maximal  $\mathbb{T}$ -isotropy subgroups do not act with the rotating circle property (Fig. 3). A possible graph for such a minimum can be found in Fig. 4, for  $q = 3$  (one needs to prove that the minimum is not the homographic solution – with a level estimate – and that there are no boundary collisions – with an argument similar to (Chen 2001a)). See also (Chen 2003a) for an updated and much generalized treatment of such orbits.

**Example 5** Consider four particles with equal masses and an even integer  $q \geq 4$ . Let be the direct product of the dihedral group of order  $2q$  with the group  $C_2$  of order 2. Let  $D_{4q}$  be presented by  $D_{4q} = \langle g_1, g_2 | g_1^2 = g_2^q = (g_1 g_2)^2 = 1 \rangle$ , and let  $c \in C_2$  be the non-trivial element of  $C_2$ . As in



**n-Body Problem and Choreographies, Fig. 3** Another symmetry constraint for an eight-shaped orbit ( $n = 5$ )



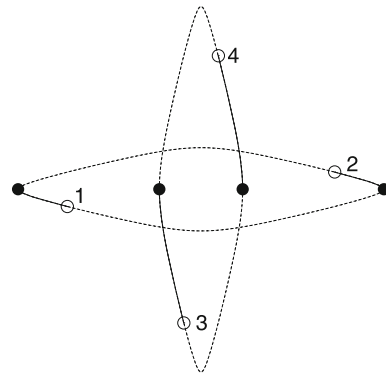
**n-Body Problem and Choreographies, Fig. 4** The orbit of Example 4 with  $q = 3$

Example 4, define the homomorphisms  $\rho, \tau$  and  $\sigma$  as follows.

$$\begin{aligned} \rho(g_1) &= \tau(g_1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ \rho(g_2) &= \tau(g_2) = \begin{bmatrix} \cos \frac{2\pi}{q} & -\sin \frac{2\pi}{q} \\ \sin \frac{2\pi}{q} & \cos \frac{2\pi}{q} \end{bmatrix}, \\ \rho(c) &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \tau(c) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \sigma(g_1) &= (1, 2)(3, 4), \sigma(g_2) = (1, 3)(2, 4), \\ \sigma(c) &= (1, 2)(3, 4). \end{aligned}$$

Again, one can show that a minimizer without interior collisions exists since  $\ker \tau = C_2$  acts with the rotating circle property (a possible minimizer is shown in Fig. 5). This generalizes Chen's orbit (Chen 2003a). See also (Chen 2001a).

**Example 6 (Hip-Hops)** If  $G = \mathbb{Z}_2$  is the group of order 2 acting trivially on  $\mathbf{n}$ , acting with the antipodal map on  $V = \mathbb{R}^3$  and on the time circle  $\mathbb{T}$ , then again  $\chi^G = 0$ , so that Proposition 2 holds. Furthermore, since the action is of cyclic type Corollary 2 ensures that minimizers have no collisions. Such minimizers were called generalized Hip-Hops in (Chenciner 2002b). See also (Chenciner 2002c). A subclass of symmetric trajectories leads to a generalization of such a Hip-Hop. Let  $n \geq 4$



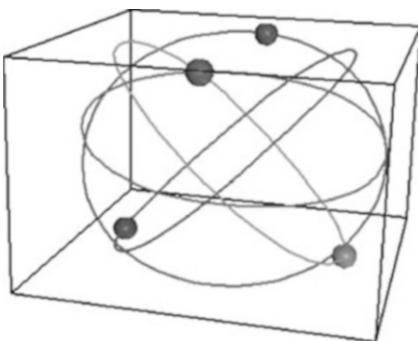
**n-Body Problem and Choreographies, Fig. 5** A possible minimizer for Example 5

an even integer. Consider  $n$  particles with equal masses, and the group  $G = C_n \times C_2$  direct product of the cyclic group of order  $n$  (with generator  $g_1$ ) and the group  $C_2$  of order 2 (with generator  $g_2$ ). Let the homomorphisms  $\rho$ ,  $\sigma$  and  $\tau$  be defined by

$$\begin{aligned} \rho(g_1) &= \begin{bmatrix} \cos \frac{2\pi}{n} & -\sin \frac{2\pi}{n} & 0 \\ \sin \frac{2\pi}{n} & \cos \frac{2\pi}{n} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \rho(g_2) &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \tau(g_1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \tau(g_2) &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \\ \sigma(g_1) &= (1, 2, 3, 4), \quad \sigma(g_2) = (). \end{aligned}$$

It is easy to see that  $X^G = 0$ , and thus a minimizer exists. Since the action is of cyclic type, it suffices to exclude interior collisions. But this follows from the fact that  $\ker \tau = C_n$  has the rotating circle property. This example is the natural generalization of the Hip-Hop solution of (Chenciner and Venturelli 2000) to  $n \geq 4$  bodies. We can see the trajectories in Fig. 6.

**Example 7** Consider the direct product  $G = D_6 \times C_3$  of the dihedral group  $D_6$  (with generators  $g_1$  and  $g_2$  of order 3 and 2 respectively) of order 6 and the cyclic group  $C_3$  of order 3 generated by  $c \in C_3$ . Let us consider the planar  $n$ -body

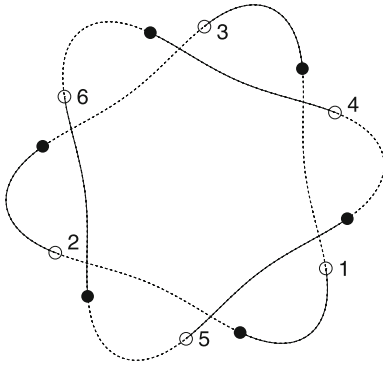


**n-Body Problem and Choreographies, Fig. 6** The Chenciner–Venturelli Hip–Hop

problem with  $n = 6$  with the symmetry constraints given by the following  $G$ -action.

$$\begin{aligned} \rho(g_1) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \rho(g_2) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \\ \rho(c) &= \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{bmatrix}, \\ \tau(g_1) &= \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{bmatrix}, \\ \tau(g_2) &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \tau(c) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \sigma(g_1) &= (1, 3, 2)(4, 5, 6), \\ \sigma(g_2) &= (1, 4)(2, 5)(3, 6), \\ \sigma(c) &= (1, 2, 3)(4, 5, 6). \end{aligned}$$

By Proposition 2 one can prove that a minimizer exists, and since  $G$  acts with the rotating circle property (actually, the elements of the image of  $\rho$  are rotations) on  $\mathbb{T}$ -maximal isotropy subgroups, the conclusion of Theorem 2 holds. It is not difficult to see that configurations in  $X^{\ker \tau}$  are given by two centered equilateral triangles. Now, to guarantee that the minimizer is not a homographic solution, of course it suffices to show that there are no homographic solutions in  $\mathcal{A}^G$  (like in the case of Example 2). This follows from the easy observation that at some times  $t \in \mathbb{T}$  with maximal isotropy it happens that  $x_1 = -x_4, x_2 = -x_5$  and  $x_3 = -x_6$ , while at some other times it happens that  $x_1 = -x_5, x_2 = -x_6$  and  $x_3 = -x_4$  or that  $x_1 = -x_6, x_2 = -x_4$  and  $x_3 = -x_5$  and this implies that there are no homographic loops in  $\mathcal{A}^G$ . With no difficulties the same action can be defined for  $n = 2k$ , where  $k$  is any odd integer. We can see a possible trajectory in Fig. 7. Also, it is not difficult to consider a similar example in dimension 3. With  $n = 6$  and the notation of  $D_6$  and  $C_3$  as above, consider the group  $G = D_6 \times C_3 \times C_2$ . Let  $g_1, g_2, c$  be as above, and let  $c_2$  be the generator of  $C_2$ . The homomorphisms  $\rho, \tau$  and  $\sigma$  are defined in a similar way by

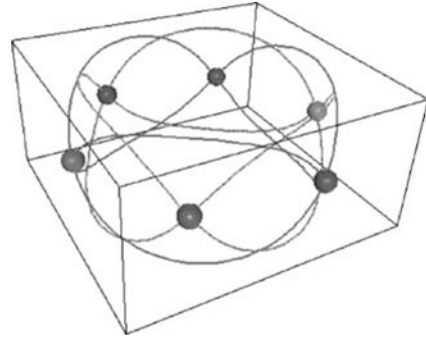


**n-Body Problem and Choreographies, Fig. 7** The planar equivariant minimizer of Example 7

$$\begin{aligned} \rho(g_1) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \rho(g_2) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \rho(c) &= \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} & 0 \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \rho(c_2) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \\ \tau(g_1) &= \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{bmatrix}, \quad \tau(g_2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ \tau(c) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \tau(c_2) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \\ \sigma(g_1) &= (1, 3, 2)(4, 5, 6), \\ \sigma(g_2) &= (1, 4)(2, 5)(3, 6), \\ \sigma(c) &= (1, 2, 3)(4, 5, 6), \quad \sigma(c_2) = (). \end{aligned}$$

In the resulting collisionless minimizer (again, it follows by Proposition 2 and Theorem 2) two equilateral triangles rotate in opposite directions and have a “brake” motion on the third axis. The likely shape of the trajectories can be found in Fig. 8.

**Example 8 (Marchal’s  $P_{12}$ -symmetry revisited)** Let  $k \geq 2$  be an integer, and consider the cyclic group  $G = C_{6k}$  of order  $6k$  generated by the element  $c \in G$ . Now consider orbits for  $n = 3$



**n-Body Problem and Choreographies, Fig. 8** The three-dimensional equivariant minimizer of Example 7

bodies in the space of dimension  $d = 3$ . With a minimal effort and suitable changes the example can be generalized for every  $n \geq 3$ . We leave the details to the reader. The homomorphisms  $\rho$ ,  $\tau$  and  $\sigma$  are defined by

$$\begin{aligned} \rho(c) &= \begin{bmatrix} \cos \frac{\pi}{k} & -\sin \frac{\pi}{k} & 0 \\ \sin \frac{\pi}{k} & \cos \frac{\pi}{k} & 0 \\ 0 & 0 & -1 \end{bmatrix}, \\ \tau(c) &= \begin{bmatrix} \cos \frac{2\pi}{6k} & -\sin \frac{2\pi}{6k} \\ \sin \frac{2\pi}{6k} & \cos \frac{2\pi}{6k} \end{bmatrix}, \\ \sigma(c) &= (1, 2, 3). \end{aligned}$$

Straightforward calculations show that  $\chi^G = 0$  and hence Proposition 2 can be applied. Furthermore, the action is of cyclic type with  $\ker \tau = 1$ , and hence by Corollary 2 the minimizer does not have collisions. It is left to show that this minimum is not a homographic motion. The only homographic motion in  $A^G$  is a Lagrange triangle  $y(t) = (y_1, y_2, y_3)(t)$ , rotating with angular velocity  $3 - 2k$  (assume that the period is  $2\pi$ , i.e. that  $T = |\mathbb{T}| = 2\pi$ ) in the plane  $u_3 = 0$  (let  $u_1, u_2, u_3$  denote the coordinates in  $\mathbb{R}^3$ ). To be a minimum it needs to be inscribed in the horizontal circle of radius  $((\alpha^3 - \alpha^2)/(2(3 - 2k)^2))^{1/(2 + \alpha)}$ . Now, for every function  $\phi(t)$  defined on  $\mathbb{T}$  such that  $\phi(c^3 t) = -\phi(t)$ , the loop given by  $v_1(t) = (0, 0, \phi(t))$ ,  $v_2(t) = (0, 0, \phi(c^{-2}t))$  and  $v_3(t) =$

$(0, 0, \phi(c^2t))$  is  $G$ -equivariant, and thus belongs to  $A^G$ . If one computes the value of Hessian of the Lagrangian action  $\mathcal{A}$  in  $y$  and in the direction of the loop  $v$  one finds that

$$D_v^2 \mathcal{A}|_y = 3 \int_0^{2\pi} \dot{\phi}^2(t) dt - 2(3 - 2k)^2 \times \int_0^{2\pi} (\phi(t) + \phi(ct))^2 dt.$$

In particular, if we set the function  $\phi(t) = \sin(kt)$ , which has the desired property, elementary integration yields

$$D_v^2 \mathcal{A}|_y = 3\pi(k^2 - 2(3 - 2k)^2),$$

which does not depend on  $\alpha$  and is negative for every  $k \geq 3$ . Thus for every  $k \geq 3$  the minimizer is not homographic. We see a possible trajectory in Fig. 9.

**Remark 1** In the previous example, if  $k \not\equiv 0 \pmod{3}$ , the cyclic group  $G$  can be written as the sum  $C_3 + C_{2k}$ . The generator of  $C_3$  acts trivially on  $V$ , acts with a rotation of order 3 on  $\mathbb{T}$  and with the cyclic permutation  $(1, 2, 3)$  on  $\{1, 2, 3\}$ . This means that for all  $k \not\equiv 0 \pmod{3}$  the orbits of Example 8 are non-planar choreographies (Fig. 10). Furthermore, it is possible to define a cyclic action of the same kind by setting  $\tau$  and  $\sigma$  as above and

$$\rho(c) = \begin{bmatrix} \cos \frac{p}{3k} \pi & -\sin \frac{p}{3k} \pi & 0 \\ [1mm] \sin \frac{p}{3k} \pi & \cos \frac{p}{3k} \pi & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

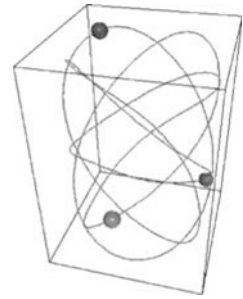
where  $p$  is a non-zero integer. If  $p = 3$  one obtains the same action as in Example 8. One can perform similar computations and obtain that the Lagrange orbit (with angular velocity  $p - 2k$ , this time) is not a minimizer for all  $(p, k)$  such that  $0 < p < 3k$  and  $k^2 - 2(p - 2k)^2 < 0$ .

### More Examples with non Trivial Core

Now consider the core of  $G$ ,  $\ker \tau = K \subset G$ . With an abuse of notation, using the factorization of the

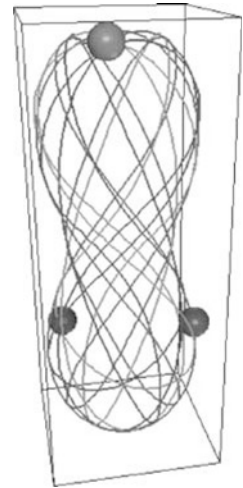
### n-Body Problem and Choreographies,

**Fig. 9** The non-planar choreography of Example 8 for  $k = 4$



### n-Body Problem and Choreographies,

**Fig. 10** A non-planar symmetric orbit of Remark 1, with  $k = 3$  and  $p = 1$



action, we can identify  $K$  with a subgroup of  $O(3)$ . A very remarkable class of symmetry groups with non trivial core has been defined by Ferrario in (Ferrario 2007), through its  $Krh$  decomposition:

**Definition 4** Let  $G$  be a symmetry group. Then define

1.  $K = \ker \tau$ ,
2.  $[r] \in W_{O(3)}K$  as the image in the Weyl group of the generator mod  $K$  of  $\ker \det \tau \subset G/K$  (corresponding to the time-shift with minimal angle). If  $\ker \det \tau = K$ , then  $[r] = 1$ .
3.  $[h] \in W_{O(3)}K$  as the image in the Weyl group of one of the time-reflections mod  $K$  in  $G/K$ , in the cases such an element exists. Otherwise it is not defined.

In short, the triple

$$(K, [r], [h])$$

is said the  $Krh$  data of  $G$ .



**Example 9** Consider the icosahedral group  $Y$  of order 60. The group  $G$  with  $Krh$  data

$$\begin{pmatrix} 1 & (1, -1) \\ Y & -1 \end{pmatrix}$$

is isomorphic to the direct product  $I \times Y$  of order 120, and acts on the euclidean space  $\mathbb{R}^3$  as the full icosahedron group. The action on  $\mathbb{T}$  is cyclic and given by the fact that  $\ker \tau = 1 \times Y$ . The isotropy is generated by the central inversion  $-1$ , and hence the set of bodies is  $G/I \cong Y$ . Thus at any time  $t$  the 60 point particles are constrained to be a  $Y$ -orbit in  $\mathbb{R}^3$  (which does not mean they are vertices of a icosahedron, simply that the configuration is  $Y$ -equivariant). After half period every body is in the antipodal position:  $x_i(t + T/2) = -x_i$  (in other words, the group contains the anti-symmetry), also known as Italian symmetry – see (Chenciner 2002b, c). Of course, the group  $Y$  is just an example: one can choose also the tetrahedral group  $T$  or the octahedral  $O$  and obtain anti-symmetric orbits for 12 (tetrahedral) or 24 (octahedral) bodies, as depicted in Fig. 11. The action is by its definition transitive and coercive; local minimizers are collisionless since the maximal  $\mathbb{T}$ -isotropy group acts as a subgroup of  $SO(3)$  (i.e. orientation-preserving).

**Example 10** Let  $G$  be the group with  $Krh$  data

$$\begin{pmatrix} 1 & (1, -1) \\ D_k & -1 \end{pmatrix},$$

where  $D_k$  is the rotation dihedral group of order  $2k$ . As in the previous example, the action is such that the action functional is coercive and its local minima collisionless. At every time instant the bodies are  $D_k$ -equivariant in  $\mathbb{R}^k$  and the anti-symmetry holds. Approximations of minima can be seen in Fig. 12.

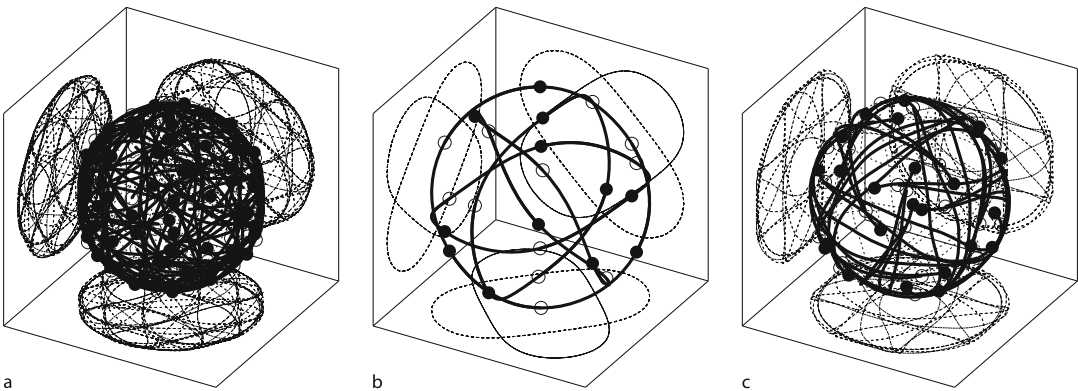
**Example 11** To illustrate the case of non-transitive symmetry group, consider the following (cyclic)  $Krh$  data:

$$\begin{pmatrix} 1 & (3, -1) \\ 1 & -1 \end{pmatrix},$$

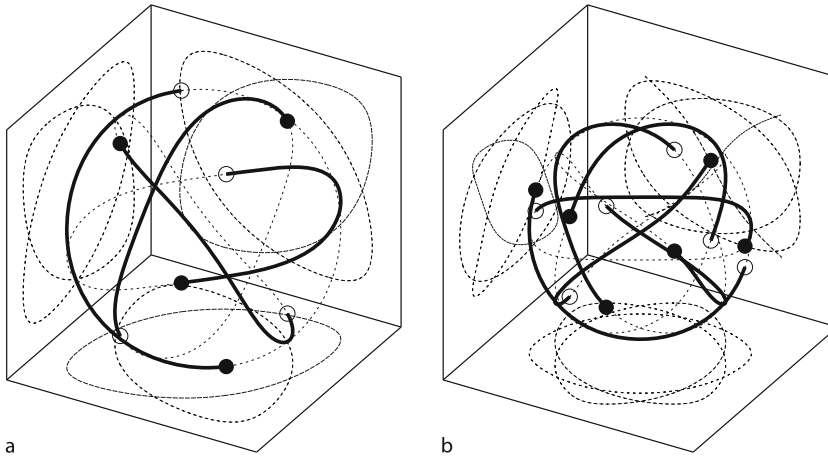
which yields a group of order 6 acting cyclically on 3 bodies, and with the antipodal map on  $\mathbb{R}^3$ . Since  $\ker \tau$  is trivial and the group is of cyclic type, local minima are collisionless. Now, by adding  $k$  copies of such group one obtains a symmetry group having  $k$  copies of it as its transitive components, where still local minimizers are collisionless and the restricted functional is coercive. Some possible minima can be found in Fig. 13, for  $k = 3, 4$ .

### The 3-Body Problem

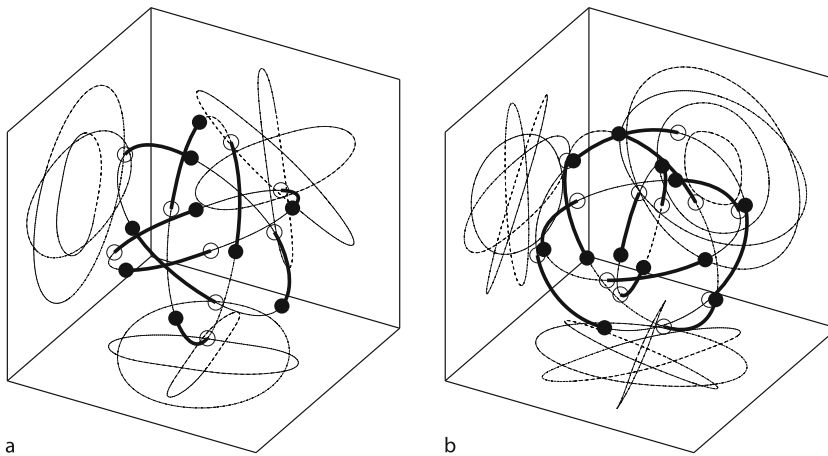
The major achievement of Barutello et al. (2008a) is to give the complete description of the outcome of the equivariant minimization procedure for the



**n-Body Problem and Choreographies, Fig. 11** 60-icosahedral  $Y$ , 12-tetrahedral  $T$  and 24-octahedral  $O$  periodic minimizers (chiral)



**n-Body Problem and Choreographies, Fig. 12** 4-dihedral  $D_2$  and 6-dihedral  $D_6$  symmetric periodic minimizers



**n-Body Problem and Choreographies, Fig. 13** 9 and 12 bodies in anti-choreographic constraints grouped by 3

planar three-body problem. First we can ensure that minimizers are always collisionless.

**Theorem 3** Let  $G$  a symmetry group of the Lagrangian in the 3-body problem (in a rotating frame or not). If  $G$  is not bound to collision (i.e. every equivariant loop has collisions), then any (possible) local minimizer is collisionless.

A symmetry group  $G$  of the Lagrangian functional  $\mathcal{A}$  is termed

- *bound to collisions* if all  $G$ -equivariant loops actually have collisions,
- *fully uncoercive* if for every possible rotation vector? the action functional  $\mathcal{A}_\omega^G$  in the frame

rotating around? with angular speed  $|\omega|$  is not coercive in the space of  $G$ -equivariant loops (that is, its global minimum escapes to infinity);

- *homographic* if all  $G$ -equivariant loops are constant up to orthogonal motions and rescaling.
- The *core* of the group  $G$  is the subgroup of all the elements which do not move the time  $t \in \mathbb{T}$ .

If, for every angular velocity,  $G$  is a symmetry group for the Lagrangian functional in the rotating frame, then we will say that  $G$  is of type  $R$ . This is a fundamental property for symmetry groups. In fact, if  $G$  is *not* of type  $R$ , it turns out that the angular momentum of all  $G$ -equivariant trajectories vanishes.

**The Classification of Planar Symmetry Groups for 3-body**

**Theorem 4** Let  $G$  be a symmetry group of the Lagrangian action functional in the planar 3-body problem. Then, up to a change of rotating frame,  $G$  is either bound to collisions, fully uncoercive, homographic, or conjugated to one of the symmetry groups listed in Table 1 (RCS stands for Rotating Circle Property and HGM for Homographic Global Minimizer) (Fig. 14).

**Planar Symmetry Groups**

- The trivial symmetry  
Let  $G$  be the trivial subgroup of order 1. It is clear that it is of type  $R$ , it has the rotating circle property. It yields a coercive functional on

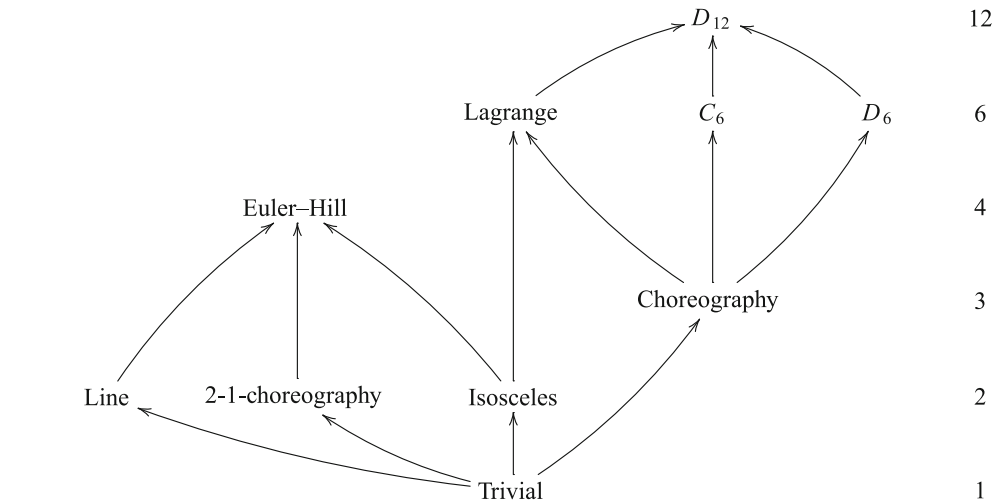
$\mathcal{A}^G = \mathcal{A}$  only when  $\omega$  is not an integer. If  $\omega = \frac{1}{2} \pmod{1}$  then the minimizers are minimizers for the anti-symmetric symmetry group (also known as *Italian* symmetry)  $x(at) = ax(t)$ , where  $a$  is the antipodal map on  $\mathbb{T}$  and  $E$ . The masses can be different.

**Proposition 3** For every  $\omega \notin \mathbb{Z}$  and every choice of masses the minimum for the trivial symmetry occurs in the relative equilibrium motion associated to the Lagrange central configuration.

- The line symmetry  
Another case of symmetry group that can be extended to rotating frames with arbitrary

**n-Body Problem and Choreographies, Table 1** Planar symmetry groups with trivial core

| Name               | $ G $ | Type $R$ | Action type | trans. dec. | RCP  | HGM |
|--------------------|-------|----------|-------------|-------------|------|-----|
| Trivial            | ?1    | yes      |             | 1 + 1 + 1   | yes  | yes |
| Line               | ?2    | yes      | brake       | 1 + 1 + 1   | (no) | no  |
| 2 – 1-choreography | ?2    | yes      | cyclic      | 2 + 1       | yes  | no  |
| Isosceles          | ?2    | yes      | brake       | 2 + 1       | no   | yes |
| Hill               | ?4    | yes      | dihedral    | 2 + 1       | no   | no  |
| 3-choreography     | ?3    | yes      | cyclic      | 3           | yes  | yes |
| Lagrange           | ?6    | yes      | dihedral    | 3           | no   | yes |
| $C_6$              | ?6    | no       | cyclic      | 3           | yes  | no  |
| $D_6$              | ?6    | no       | dihedral    | 3           | yes  | no  |
| $D_{12}$           | 12    | no       | dihedral    | 3           | no   | no  |



**n-Body Problem and Choreographies, Fig. 14** The poset of symmetry groups for the planar 3-body problem

masses is the line symmetry: the group is a group of order 2 acting by a reflection on the time circle  $\mathbb{T}$ , by a reflection on the plane  $E$ , and trivially on the set of indexes. That means, at time 0 and  $\pi$  the masses are collinear, on a fixed line  $l \subset E$ . It is coercive only when  $\omega \in \mathbb{Z}$ . In this case the Lagrangian solution cannot be a minimum, while it can be the relative equilibrium associated with the Euler configuration.

- *The 2 – 1-choreography symmetry*

Consider the group of order 2 acting as follows:  $\rho(g) = 1$ ,  $\tau(g) = -1$  (that is, the translation of half-period) and  $\sigma(g) = (1, 2)$  (that is,  $\sigma(g)(1) = 2$ ,  $\sigma(g)(2) = 1$ , and  $\sigma(g)(3) = 3$ ). That is, it is a half-period choreography for the bodies 1 and 2. It can be extended to rotating frames and coercive for a suitable choice of  $\omega \neq 0, 1 \pmod 2$ .

The Euler's orbit with  $k = 1$  and the Hill's orbits with  $k = \pm 1$  are equivariant for the 2 – 1-choreography symmetry, while the Euler's orbit with  $k = 0$  is not equivariant for this symmetry. In Figs. 15 and 16 Euler 1 represents the action levels on the Euler's orbit with  $k = 1$ , Hill 1–2 the ones on the Hill's with  $k = \pm 1$ .

- *The isosceles symmetry*

The isosceles symmetry can be obtained as follows: the group is of order 2, generated by  $h$ ;  $\tau(h)$  is a reflection in the time circle  $\mathbb{T}$ ,  $\rho(h)$  is a reflection along a line  $l$  in  $E$ , and  $\sigma(h) = (1, 2)$  as above. The constraint is therefore that at time 0 and  $\pi$  the 3-body configuration is an isosceles triangle with one vertex on 1 (the third).

**Proposition 4** For every  $\omega \notin \mathbb{Z}$  and every choice of masses the minimum for the isosceles symmetry occurs in the relative equilibrium motion associated to the Lagrange configuration.

- *The Euler–Hill symmetry*

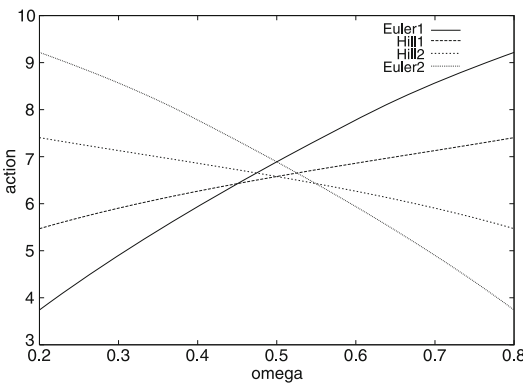
Now consider the symmetry group with a cyclic generator  $r$  of order 2 (i.e.  $\tau(r) = -1$ ) and a time reflection  $h$  (i.e.  $\tau(h)$  is a reflection of  $\mathbb{T}$ ) given by  $\rho(r) = 1$ ,  $\sigma(r) = (1, 2)$ ,  $\rho(h)$  is a reflection and  $\sigma(h) = ()$ . It contains the 2 – 1-choreography (as the subgroup  $\ker \det(\tau)$ ), the isosceles symmetry (as the isotropy of  $\pi/2 \in \mathbb{T}$ ) and the line symmetry (as the isotropy of  $0 \in \mathbb{T}$ ) as subgroups.

**Proposition 5** The minimum of the Euler–Hill symmetry is not homographic, provided that the angular velocity  $\omega$  is close to 0.5 and the values of the masses are close to 1.

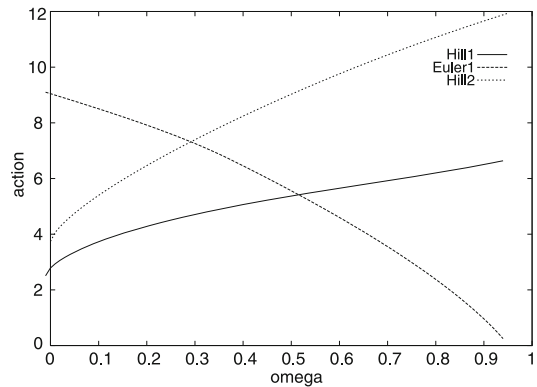
- *The choreography symmetry*

The choreography symmetry is given by the group  $C_3$  of order 3 acting trivially on the plane  $E$ , by a rotation of order 3 in the time circle  $\mathbb{T}$  and by the cyclic permutation  $(1, 2, 3)$  of the indices.

**Proposition 6** For every  $\omega$  the minimal choreography of the 3 body problem is a rotating Lagrange configuration.



**n-Body Problem and Choreographies, Fig. 15** Action levels for the line symmetry



**n-Body Problem and Choreographies, Fig. 16** Action levels for the 2 – 1-choreography symmetry

- The Lagrange symmetry

The Lagrange symmetry group is the extension of the choreography symmetry group by the isosceles symmetry group. Thus, it is a dihedral group of order 6, the action is of type  $R$ . Hence, the relative equilibrium motions associated to the Lagrange configuration are admissible motions for this symmetry and, again, the minimizer occurs in the relative equilibrium motion associated to the Lagrange configuration.

- The Chenciner–Montgomery symmetry group and the eights

There are three symmetry groups (up to change of coordinates) that yield the Chenciner–Montgomery *figure eight* orbit: they are the only symmetry groups which do not extend to the rotating frame and we have already described them in subsection “[The Eight Shaped Three-Body Solution](#)”. One can prove that all  $G$ -equivariant trajectories have vanishing angular momentum, whenever the group is not of type  $R$ . Moreover, we were able to partially answer to the open question (posed by Chenciner) whether their minimizers coincide or not: for two of them ( $D_6$  and  $D_{12}$ ) the minimizer is necessarily the same.

### Space Three-body Problem

Based on the classification of planar groups, by introducing a natural notion of space extension of a planar group, Ferrario gave in (Ferrario 2006) a complete answer to the classification problem for the three-body problem in the space and at the same time to determine the resulting minimizers and describe its more relevant properties.

**Theorem 5** Symmetry groups not bound to collisions, not fully uncoercive and not homographic are, up to a change of rotating frame, either the three-dimensional extensions of planar groups (if trivial core) listed in Table 2 below or the vertical isosceles triangle (if non-trivial core).

The next theorem is the answer to the natural questions about collisions and description of some main features of minimizers.

**n-Body Problem and Choreographies, Table 2** Space extensions of planar symmetry groups with trivial core

| Name           | Extensions                        |
|----------------|-----------------------------------|
| Trivial        | $C_1^-$                           |
| Line           | $L_2^{+,-}, L_2^{-,+}$            |
| Isosceles      | $H_2^{+,-}, H_2^{-,+}$            |
| Hill           | $H_4^{+,-}, H_4^{-,+}$            |
| 3-choreography | $C_3^+, C_3^-$                    |
| Lagrange       | $L_6^{+,+}, L_6^{+,-}, L_6^{-,+}$ |
| $D_6$          | $D_6^{+,-}, D_6^{-,+}$            |
| $D_{12}$       | $D_{12}^{+,-}$                    |

**Theorem 6** Let  $G$  be a symmetry group not bound to collisions and not fully uncoercive. Then

1. Local minima of  $\mathcal{A}^\omega$  do not have collisions.
2. In the following cases minimizers are planar trajectories:
  - (a) If  $G$  is not of type  $R$ :  $D_6^{+,-}, D_6^{-,+}$  and  $D_{12}^{+,-}$  (and then  $G$ -equivariant minimizers are Chenciner–Montgomery eights).
  - (b) If there is a  $G$ -equivariant minimal Lagrange rotating solution:  $C_1^-, H_2^{+,-}, C_3^+, L_6^{+,+}, L_6^{+,-}$  (and then the Lagrange solution is of course the minimizer).
  - (c) If the core is non-trivial and it is not the vertical isosceles (and then minimizers are homographic).
3. In the following cases minimizers are always non-planar:
  - (a) The groups  $L_6^{-,+}$  and  $C_3^-$  for all  $\omega \in (-1, 1) + 6\mathbb{Z}$ ,  $\omega \neq 0$  (the minimizers for  $L_6^{-,+}$  are the elements of Marchal family  $P_{12}$ , and minimizers of  $C_3^-$  are a less-symmetric family  $P'_{12}$ . (Highly likely they are not distinct families: this is the recurring phenomenon of “more symmetries than expected” in n-body problems.)
  - (b) The extensions of line and Hill–Euler type groups, for on open subset of mass distributions and angular speeds?:  $L_2^{+,-}, L_2^{-,+}, H_4^{+,-}$  and  $H_4^{-,+}$  (for  $L_2^{+,-}$  this happens also with equal masses).

- (c) The vertical isosceles for suitable choices of masses and  $\omega$ .

## Minimizing Properties of Simple Choreographies

In the space of symmetric (choreographical) loops, the action takes the form

$$\begin{aligned} \mathcal{A}(x) = & \frac{1}{2} \sum_{h=0}^{n-1} \int_0^\tau |\dot{x}(t + h\tau)|^2 dt + \frac{1}{2} \\ & \times \sum_{\substack{h,l=0 \\ h \neq l}}^{n-1} \int_0^\tau \frac{dt}{|x(t + l\tau) - x(t + h\tau)|^\alpha}. \end{aligned}$$

A natural question concerns the nature of the minimizers under the sole choreographical constraint. Unfortunately, the bare minimization among choreographical loops gives returns only trivial motions:

**Theorem 7** For every  $\alpha \in \mathbb{R}_+^*$  and  $d \geq 2$ , the absolute minimum of  $\mathcal{A}$  on  $\mathcal{A}$  is attained on a relative equilibrium motion associated to the regular  $n$ -g on.

This theorem extends some related result for the Italian symmetry by Chenciner and Desolneux (1998), and the results in section “The 3-Body Problem”. The proof is based on a (quite involved) convexity argument together with the analysis of some spectral properties related to the choreographical constraint. Now, in order to find nontrivial minimizers, we look at the same problem in a rotating frame. In order to take into account of the Coriolis force, the new action functional has to contain a gyroscopic term:

$$\begin{aligned} \mathcal{A}(y) = & \frac{1}{2} \int_0^{2\pi} |\dot{y}(t) + J\omega y(t)|^2 dt \\ & + \frac{1}{2} \sum_{h=1}^{n-1} \int_0^{2\pi} \frac{dt}{|y(t) - y(t + h\tau)|^\alpha}. \end{aligned}$$

Consider the function  $h : \mathbb{R}_+^* \rightarrow \mathbb{N}$ ,  $h(\omega) = \min_{n \in \mathbb{N}^*} (\omega - n)^2 / n^2$  and let  $\omega^* = \frac{4}{3}$ . The same technique used for the inertial system extends to

rotating systems having small angular velocity; this gives the following result (Figs. 17 and 18).

**Theorem 8** If  $\omega \in (0, \omega^*) \setminus \{1\}$ , then the action attains its minimum on a circle with minimal period  $2\pi$  and radius depending on  $n$ ,  $\alpha$  and  $\omega$ .

### When $\omega$ is Close to an Integer

The situation changes dramatically when  $\omega$  is close to some integer. To understand this phenomenon, let us first check the result of the minimization procedure when  $\omega$  is an integer:

### Proposition 7

|     |                                                                                                                          |
|-----|--------------------------------------------------------------------------------------------------------------------------|
| (1) | If $\omega = n$ , then the action has a continuum of minimizers.                                                         |
| (2) | If $\omega = k$ , coprime with $n$ , then the action does not achieve its infimum (escape of minimizing sequences).      |
| (3) | If $\omega = k$ and $k$ divides $n$ , then the action does not achieve its infimum (clustering of minimizing sequences). |

As a consequence, we have the following result:

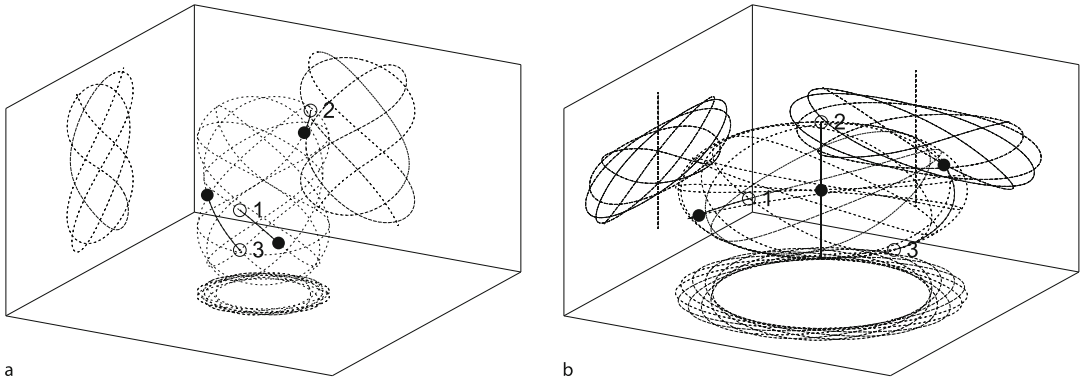
**Theorem 9** Suppose that  $n$  and  $k$  are coprime. Then there exist  $\epsilon = \epsilon(\alpha, n, k)$  such that if  $\omega \in (k - \epsilon, k + \epsilon)$  the minimum of the action is attained on a circle with minimal period  $2\pi/k$  that lies in the rotating plane with radius depending on  $n$ ,  $\alpha$  and  $\omega$ .

An interesting situation appears when the integer closest to the angular velocity is not coprime with the number of bodies. In this case we prove that the minimal orbit it is not circle anymore, as the following theorem states.

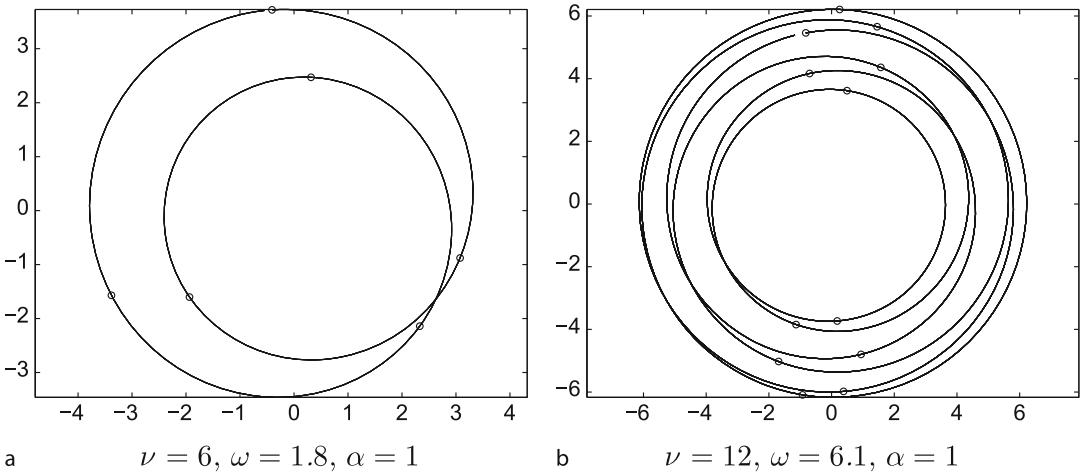
**Theorem 10** Take  $k \in \mathbb{N}$  and  $g.c.d.(k, n) = \tilde{k} > 1$ ,  $\tilde{k} \neq n$ . Then there exists  $\epsilon = \epsilon(\alpha, n, k) > 0$  such that if  $\omega \in (k - \epsilon, k + \epsilon) \setminus \{k\}$  the minimum of the action is attained on a planar  $2\pi$ -periodic orbit with winding number  $k$  which is not a relative equilibrium motion.

Also, it has be noticed that, for large number of bodies and angular velocities close to the half on an integer, the minimizer apparently is not anymore planar.





**n-Body Problem and Choreographies, Fig. 17** Non planar minimizers for the groups  $L_6^{-,+}$  and  $L_2^{+,-}$ , plotted in the fixed frame

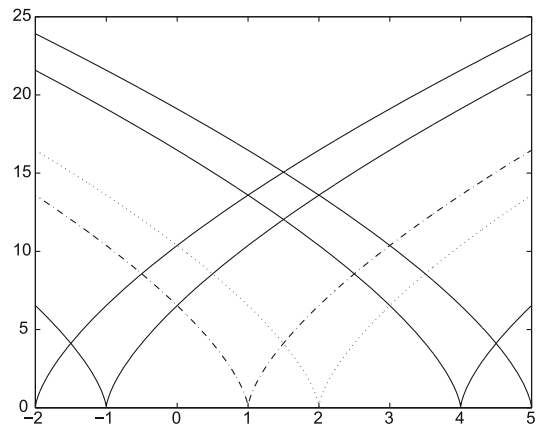


**n-Body Problem and Choreographies, Fig. 18** Examples for Theorem 10 close to an integer that divides  $n$

### Mountain Pass Solutions for the Choreographical 3-Body Problem

The discussion carried in the previous section shows that, as the angular velocity varies, the minimizer's shape must undergoes some transitions (for example it has to pass from relative equilibrium having different winding numbers). This scenario suggests the presence of other critical points, such as local minimizers or mountain pass. This was indeed discovered numerically in Barutello and Terracini (2004) and then proved by a computer assisted proof in Arioli et al. (2006).

To begin with, let us look at the following figure where the values of the action functional  $\mathcal{A}^\omega$  on the branches of circular orbits  $L_k^\omega$  are plotted:



The analysis of this picture suggests the presence of critical points different from the Lagrange motions. Indeed, let us take the angular velocity

$\omega = 1.5$ : in this case there are two distinct global minimizers, the uniform circular motions with minimal period  $2\pi$  and  $\pi$ , lying in the plane orthogonal to the rotation direction. This is a well known structure in Critical Point Theory, referred as the Mountain Pass geometry and gives the existence of a third critical point, provided the Palais–Smale condition is fulfilled, with an additional information on the Morse index. Next theorem follows from the application of the Mountain Pass Theorem to the action functional  $\mathcal{A}^{3/2}$ :

**Theorem 11** There exists a (possibly collision) critical point for the action functional  $\mathcal{A}^{3/2}$  with Morse index smaller than 1 and distinct from any Lagrange motion.

Once the existence of a Mountain Pass critical point was theoretically established, we studied its main properties in order to understand whether it belonged to some known families of periodic trajectories. To this aim we applied the bisection algorithm proposed in Barutello and Terracini (2007) to approximate the maximal of a locally optimal path joining the two strict global minimizers, finding in this a good numerical candidate. Of course, there could be a gap between the mountain pass solution whose existence is ensured by Theorem 11 and the numerical candidate found by applying the bisection algorithm. In order to fill this gap proved the existence of an actual solution very close to the numerical output of the Mountain Pass algorithm. The argument was based upon a fixed point principle and involved a rigorous computer assisted proof. As a consequence, we obtained the existence of a new branch of solution for the spatial 3-body problem (see Fig. 8).

Here are some relevant features of the new solution: the orbit is not planar, its winding number with respect, for instance, to the line  $x = -0.2$ ,  $y = 0$  is 2 and it does not intersect itself. A natural question is whether this solution can be continued as a function of the parameter?. We were able to extend the numerical–rigorous argument to cover a full interval of values of the angular velocity, providing the existence of a full branch of solution.

**Theorem 12** There exists a smooth map  $B(\omega)$  giving the (locally unique, up to symmetries)

branch of solution of the choreographical 3-body problem for all  $\omega \in [1, 2]$ , starting at the Mountain Pass solutions for  $\omega = 1.5$ .

A natural question is whether this mountain pass branch meets one of the known branches of choreographical periodic orbits: either one of the Lagrange or Marchal’s  $P_{12}$  (described by Marchal in (Marchal 2000)) families. Surprisingly enough, it turns out from further numerical computation that this branch does not emanates from any of Lagrange motions  $L_1$  or  $L_2$ ; apparently, it emanates from the branch of  $P_{12}$  solutions. The details of the bifurcation diagram are depicted in Fig. 8.

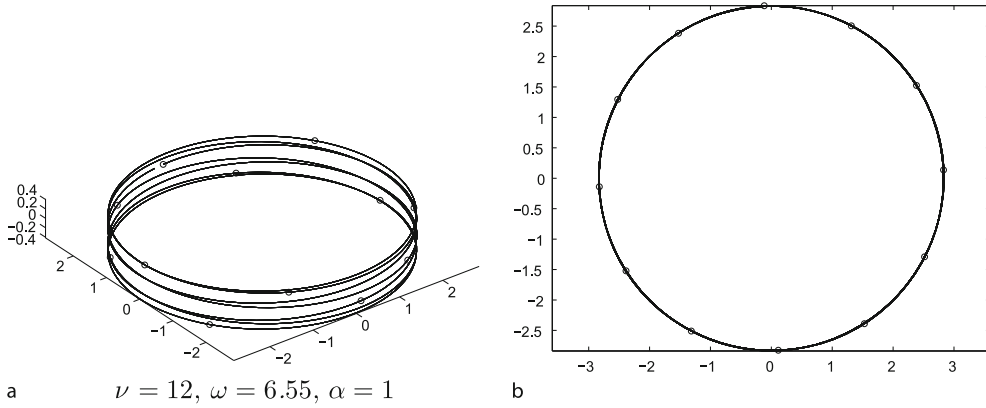
## Generalized Orbits and Singularities

The existence of a  $G$ -equivariant minimizer of the action is a simple consequence of the direct method of the Calculus of Variations. These trajectories, however, solve the associated differential equations only where they are far from the singular set. Indeed, generally speaking,  $G$ -equivariant minimizers may present singularities, even though the set of collision instants must clearly be of vanishing Lebesgue measure. A first natural question concerns the number of possible collision instants. Here below we follow (Barutello et al. 2008b) (Figs. 19, 20, and 21).

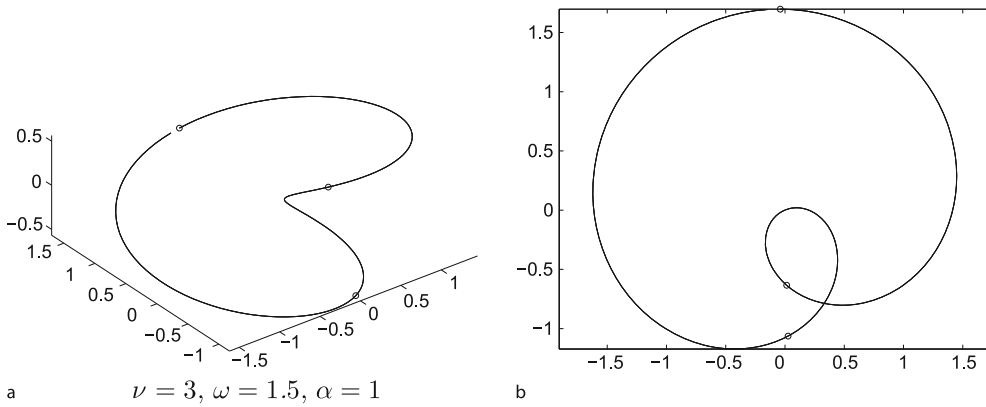
**Definition 5** A path  $x : (a, b) \rightarrow \mathcal{X}$  is called a locally minimizing solution of the N-body problem if, for every  $t_0 \in (a, b)$ , there exists  $\delta > 0$  such that the restriction of  $x$  to  $[t_0 - \delta, t_0 + \delta]$  is a local minimizer for the action with the same boundary conditions. A path which is the uniform limit of a sequence of locally minimal solutions will be called a generalized solution.

We remark that:

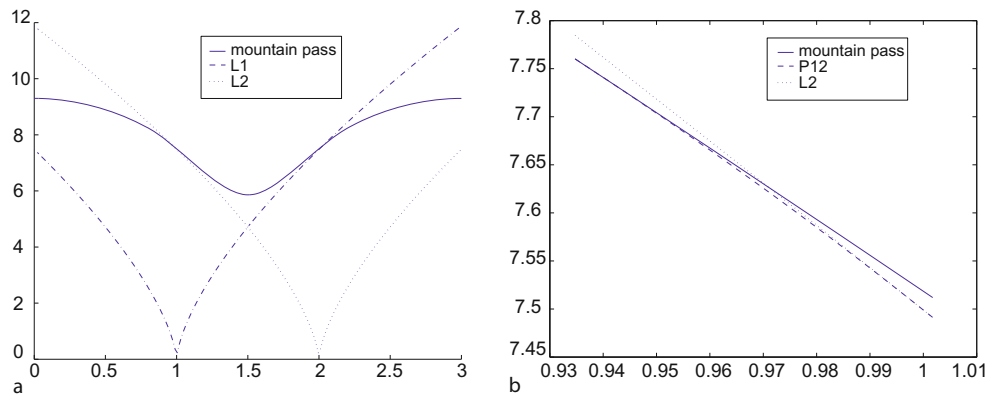
- The definition can be modified in a obvious manner to include symmetries.
- Equivariant minimizers are generalized solutions.
- If the potential is of class  $C^2$  outside collisions then every non collision solution is a generalized solution.



**n-Body Problem and Choreographies, Fig. 19** Non planar minimizers of the action with angular velocities close to the half on an integer



**n-Body Problem and Choreographies, Fig. 20** Mountain pass solution with angular velocities close to the half on an integer



**n-Body Problem and Choreographies, Fig. 21** Action levels for the Lagrange and the mountain pass solution in the 3-body problem. On the  $x$ -axes the angular velocity

varies in the interval  $[0, 3)$ . The *left picture* focuses on the bifurcation from the  $P_{12}$  family

- The generalized solutions possess an index: the minimal number of intervals  $I_j$  needed to cover  $(a, b)$  such that the restriction of  $x$  to  $I_j$  is a local minimizer for the action.
- There is a natural notion of maximal existence interval for generalized solutions, even without the unique extension property.

### Singularities and Collisions

Generally speaking we are dealing with systems of the form

$$M\ddot{x} = \nabla U(t, x), \quad t \in (a, b), \quad M_{ij} = m_i \delta_{ij},$$

where the potential  $U$  possesses a singular set  $(a, b) \times \Delta$ , in the sense that

$$[U0] \quad \lim_{x \rightarrow \Delta} U(t, x) = +\infty, \quad \text{uniformly in } t.$$

The set  $\Delta$  is clearly the *collision set*. We assume  $\Delta$  to be a *cone* in  $\mathbb{R}^{nd}$ :

$$x \in \Delta \Rightarrow \lambda x \in \Delta.$$

**Definition 6** We say that a generalized solution  $x$  on the interval  $(a, b)$ , has a *singularity* at  $t^* < +\infty$  if

$$\limsup_{t \rightarrow t^*} U(t, x(t)) = +\infty.$$

When  $t^* \in (a, b)$  we will say that  $x$  has an *interior singularity* at  $t = t^*$ , while when  $t^* = a$  or  $t^* = b$  (when finite) we will talk about a *boundary singularity*.

### The Theorems of Painlevé and Von Zeipel

In the usual mathematical language a classical solution  $x$  on the interval  $(a, b)$ , has a singularity at  $t^* < +\infty$  if it is not possible to extend  $x$  as a (classical) solution to a larger interval  $(a, t^* + \delta)$ . A classical results relates singularities of solutions with those of the potential.

**Theorem 13 (Painlevé's Theorem)** Let  $\bar{x}$  be a classical solution for the  $n$ -body dynamical system on the interval  $[0, t^*)$ . If  $\bar{x}$  has a singularity at  $t^* < +\infty$ , then

$$\lim_{t \rightarrow t^*} U(\bar{x}(t)) = +\infty.$$

Painlevé's Theorem does not necessarily imply that a *collision* (i.e. that is a singularity such the configuration has a definite limit) occurs when there is a singularity at a finite time (on this subject we refer to (McGehee 1986; Pollard and Saari 1968, 1970)). The next result has been stated by Von Zeipel in 1908 and definitely proved by Sperling in 1970.

**Theorem 14 (Von Zeipel's Theorem)** If  $\bar{x}$  is a classical solution for the  $n$ -body dynamical system on the interval  $(a, t^*)$  with a singularity at  $t^* < +\infty$  and

$$\lim_{t \rightarrow t^*} \|\bar{x}(t)\| < +\infty,$$

then  $\bar{x}(t)$  has a definite limit configuration  $x^*$  as  $t$  tends to  $t^*$ .

### Von Zeipel's Theorem and the Structure of the Collision Set

To the aim of extending the Von Zeipel's Theorem to our notion of generalized solutions, we need to introduce some assumptions on the potential  $U$  and its singular set  $\Delta$ .

$$\Delta = \bigcup_{\mu \in \mathcal{M}} V_\mu, \quad (5)$$

where the  $V_\mu$ 's are distinct linear subspaces of  $\mathbb{R}^k$  and  $\mathcal{M}$  is a finite set; observe that the set  $\Delta$  is a cone, as required before. We endow the family of the  $V_\mu$ 's with the inclusion partial ordering and we assume the family to be closed with respect to intersection (thus we are assuming that  $\mathcal{M}$  is a semi lattice of linear subspaces of  $\mathbb{R}^k$ : it is the intersection semilattice generated by the arrangement of maximal subspaces  $V_\mu$ 's). With each  $\xi \in \Delta$  we associate

$$\begin{aligned} \mu(\xi) &= \min \{ \mu : \xi \in V_\mu \} \quad \text{i.e.,} \quad V_{\mu(\xi)} \\ &= \bigcap_{\xi \in V_\mu} V_\mu. \end{aligned}$$

Fixed  $\mu \in \mathcal{M}$  we define the set of collision configurations satisfying

$$\Delta_\mu = \{\xi \in \Delta : \mu(\xi) = \mu\}$$

and we observe that this is an open subset of  $V_\mu$  and its closure  $\overline{\Delta_\mu}$  is  $V_\mu$ . We also notice that the map  $\xi \rightarrow \dim(V_{\mu(\xi)})$  is lower semi continuous.

We denote by  $p_\mu$  the orthogonal projection onto  $V_\mu$  and we write

$$x = p_\mu(x) + w_\mu(x),$$

where, of course,  $w_\mu = \mathbb{I} - p_\mu$ .

We assume that, near the collision set, the potential depends, roughly, only on the projection orthogonal to the collision set: more precisely we assume

[U5] For every  $\xi \in \Delta$ , there is  $\varepsilon > 0$  such that  
 $U(t, x) - U(t, w_{\mu(\xi)}(x)) = W(t, x)$   
 $\in C^1((a, b) \times B_\varepsilon(\xi))$ ,  
 where  $B_\varepsilon(\xi) = \{x : |x - \xi| < \varepsilon\}$ .

With these assumptions, we can extend then Von Zeipel's Theorem to generalized solutions:

**Theorem 15** Let  $\bar{x}$  be a generalized solution on the bounded interval  $(a, b)$ . If  $\bar{x}$  is bounded on the whole interval  $(a, b)$  then the singularities of  $\bar{x}$  are collisions.

## Asymptotic Estimates at Collisions

### One Side Conditions on the Potential and Its Radial Derivative

From now on, we require our potential  $U$  to satisfy some *one-side homogeneity* conditions:

[U1] there exists  $C_1 \geq 0$  such that for every  $(t, x) \in ((a, b) \times \mathbb{R}^{nd} \setminus \Delta)$

$$\left| \frac{\partial U}{\partial t}(t, x) \right| \leq C_1 U(t, x).$$

[U2] there exist  $\alpha \in (0, 2)$ ,  $\gamma > 0$  and  $C_2 \geq 0$  such that

$$\nabla U(t, x) \cdot x + \alpha U(t, x) \geq -C_2 |x|^\gamma U(t, x),$$

whenever  $|x|$  is small.

We observe that when  $U$  is homogeneous functions of degree  $-\alpha$ , the equality in condition (U2) is attained with  $C_2 = 0$ ; this assumption is satisfied also when we take into account potentials of the form  $U(t, x) = U_\alpha(x) + U_\beta(x)$  where  $U_\alpha$  is homogeneous of degree  $-\alpha$ ,  $U_\beta$  is homogeneous of degree  $-\beta$ ,  $0 < \beta < \alpha$ , and  $U_\alpha$  is positive.

### Isolatedness of Collisions Instants

In order to study the behavior of the solution at a collision, we first perform a translation in such a way that the collision holds at the origin; next we introduce *polar coordinates*:

$$r = \sqrt{Mx \cdot x} = \sqrt{I}, \quad s = \frac{x}{r},$$

where  $s \in \mathcal{E} = \{x : I^2(x) = Mx \cdot x = 1\}$  belongs to the ellipsoid of all the configurations having unitary moment of inertia.

[U3<sub>h</sub>] there exists a function  $\tilde{U}$  defined on  $(a, b) \times (\mathcal{E} \setminus \Delta)$ , such that (on compact subsets of  $((a, b) \times \mathcal{E} \setminus \Delta)$ ):

$$\lim_{r \rightarrow 0} r^\alpha U(t, rs) = \tilde{U}(t, s);$$

[U3<sub>l</sub>] there exists a function  $\tilde{U}$  defined on  $(a, b) \times (\mathcal{E} \setminus \Delta)$ , such that (on compact subsets of  $((a, b) \times \mathcal{E} \setminus \Delta)$ ):

$$\lim_{|x| \rightarrow 0} [U(t, x) + M(t) \log |x|] = \tilde{U}(t, s);$$

The following regularity theorem is proved in Barutello et al. (2008b) based on a suitable variant of Sundman's inequality and the asymptotic analysis of possible collisions outlined in the sequel.

**Theorem 16** Let  $x : (a, b) \rightarrow \mathcal{X}$  be a generalized solution of the  $N$ -body problem, with a potential  $U$  satisfying [U0–U3]. Then collision instants are isolated in  $(a, b)$ . Furthermore, if  $(a, b)$  is the finite maximal extension interval of  $x$  and no escape in finite time occurs then the number of collision instants is finite.

Some remarks are in order:

- A generalized solution does not solve the Euler–Lagrange equation in a distributional

sense (the force field can not be locally integrable). Moreover its action needs not to be finite: one needs to prove it.

- a priori our solution can have a huge set of collision instants (more than countable, but null measure);
- there may be accumulation of partial collisions at a collision having more bodies;
- there is no a priori bound on the total action on the whole interval  $(a, b)$ , nor on the energy;
- we have very weak assumptions on the potential and only a one side inequality on the radial component
- the theorem extends to the logarithmic potentials.

#### Conservation Laws

Even though they can not satisfy the differential equations in a distributional sense, generalized solutions satisfy a number of *conservation laws*.

**Theorem 17** Let  $x$  be a generalized solutions on  $(a, b)$ . Then

- The action  $\mathcal{A}(x, [a, b])$  on  $(a, b)$  is finite
- The energy  $h$  is bounded and belongs to the Sobolev space  $W^{1,1}((a, b), \mathbb{R})$ .
- Lagrange–Jacobi inequality holds in the sense of measures:

$$\frac{1}{2} \dot{I}(\bar{x}(t)) \geq 2h(t)(\bar{x}(t)) + (2 - \alpha)U(t, \bar{x}(t)) - C_2 |\bar{x}|^\gamma U(t, \bar{x}), \quad \forall t \in (a, b).$$

- A monotonicity formula holds (extending Sundman's inequality): let us consider the angular energy

$$\Gamma_\alpha(r, s) := r^\alpha \left[ \frac{1}{2} r^2 |\dot{s}|^2 - U(t, rs) \right]$$

the  $\Gamma_\alpha$  is bounded variation and

$$\begin{aligned} \frac{d}{dt} \Gamma_\alpha(r, s) &\geq -\frac{2-\alpha}{2} r^{1+\alpha} \dot{r} |\dot{s}|^2 - C_1 r^\alpha U(t, rs) \\ &\quad + C_2 r^{\alpha+\gamma} \frac{\dot{r}}{r} U(t, rs). \end{aligned}$$

#### Generalized Sundman–Sperling Estimates

The asymptotic analysis along a single collision (total or partial) trajectory goes back, in the classical case, to the works by Sundman (1913), Wintner (1941) and, in more recent years by Sperling, Pollard, Saari, Diacu and other authors (see for instance (Diacu 1992; ElBialy 1990; Pollard and Saari 1968, 1970; Saari 1972; Sperling 1970)). Now we are in a position to extend such estimate also to generalized solutions.

#### Theorem 18

- The following asymptotic estimates hold:

$$\begin{aligned} r &\sim (\kappa t)^{\frac{2}{2+\alpha}} \quad \left( r \sim |t| \sqrt{-\log(|t|)} \right) \\ K \sim U &\sim \frac{1}{4-2\alpha} \ddot{I} \sim \frac{2}{(2+\alpha)^2} \kappa^2 (\kappa t)^{\frac{-2\alpha}{2+\alpha}} \\ &\quad (\sim -\log|t|). \end{aligned}$$

- Let  $s$  be the *normalized configuration* of the colliding cluster  $s = x/r$ . Then

$$\begin{aligned} \lim_{t \rightarrow 0} r^2 |\dot{s}|^2 &= 0 \\ \lim_{t \rightarrow 0} U(s(t)) &= b < +\infty \end{aligned}$$

- Moreover there exists the angular blow-up, that is angular scaled family  $(s(\lambda t))_\lambda$  is pre-compact for the topology of uniform convergence on compact sets of  $\mathbb{R} \setminus \{0\}$ .

As a further consequence we have the *vanishing of the total angular momentum*.

In the same setting of the Theorem, assume that the potential  $U$  verifies the further assumption:

$$[\text{U4}] \quad \lim_{r \rightarrow 0} r^{\alpha+1} \nabla_T U(t, x) = \nabla_T \tilde{U}(t, s).$$

Then we have

$$\lim_{t \rightarrow t_0} \text{dist}(C^b, s(t)) = \lim_{t \rightarrow t_0} \inf_{\bar{s} \in C^b} |s(t) - \bar{s}| = 0,$$

where  $C^b$  is the set of central configurations for  $\tilde{U}$  at level  $b$ .



### Dissipation and McGehee Coordinates

The main tool in proving the asymptotic estimates follow is the *monotonicity formula*, which is somehow equivalent to *Sundman's inequality*. To fix our mind, let us think to a homogeneous potential  $U_\alpha$ . A possible way to see this dissipation is to perform the change of variables (reminiscent of the McGehee coordinates):

$$\rho = r^{\frac{2+\alpha}{4}}, \quad \rho' = \frac{2-\alpha}{4} r^{-\frac{2+\alpha}{4}} r'$$

to obtain the action functional depending on  $(\rho, s)$ :

$$\begin{aligned} \mathcal{A}(\rho, s) = & \int_0^{\tau^*} \frac{1}{2} \left( \frac{4}{2-\alpha} \right)^2 (\rho')^2 \\ & + \rho^2 \left( \frac{1}{2} |s'|^2 + U_\alpha(s) \right) - \lambda \rho^\beta d\tau, \end{aligned}$$

where  $\beta := 2(2+\alpha)/(2-\alpha) > 2$ . Here we have reparametrized the time as

$$dt = r^{\frac{2+\alpha}{2}} d\tau,$$

one proves that

$$\tau^* = \int_0^1 r^{-\frac{2+\alpha}{2}} dt = +\infty.$$

This is the coupling between a Duffing equation and the  $N$ -body angular system. When  $\rho$  is increasing, it acts as a viscosity coefficient on the angular Lagrangian.

A classical framework for the study of collisions is given by the McGehee coordinates (McGehee 1974) (here and below we assume, for simplicity of notations, all the masses be equal to one):

$$\begin{aligned} r &= |x| = I^{1/2} \\ s &= \frac{x}{r} \\ v &= r^{\alpha/2} (y \cdot s) \\ u &= r^{\alpha/2} (y - y \cdot ss). \end{aligned}$$

The equation of motions become (here,  $\dot{\phantom{x}}$  denotes differentiation with respect to the new time variable  $\tau$ ):

$$\begin{aligned} r' &= rv \\ v' &= \frac{\alpha}{2} v^2 + |u|^2 - \alpha U_\alpha(s) \\ s' &= u \\ u' &= \left( \frac{\alpha}{2} - 1 \right) vu - |u|^2 s + \alpha U_\alpha(s) s + \nabla U_\alpha(s). \end{aligned}$$

It is worthwhile noticing that the three last equations do not depend on  $r$  and hence are defined also for  $r = 0$ . In this context, the monotonicity formula means that  $v$  is a Lyapunov function for the system.

### Blow-ups

For every  $\lambda > 0$  let

$$x^\lambda(t) = \lambda^{-2/(2+\alpha)} x(\lambda t).$$

If  $\{\lambda_n\}_n$  is a sequence of positive real numbers such that  $s(\lambda_n)$  converges to a normalized configuration  $\bar{s}$ , then

$$\forall t \in (0, 1) : \lim_{n \rightarrow \infty} s(\lambda_n t) = \lim_{n \rightarrow \infty} s(\lambda_n) = \bar{s}.$$

Hence the rescaled sequence will converge uniformly to the *blow-up* of  $x(t)$  relative to the colliding cluster  $\mathbf{k} \subset \mathbf{n}$  (in  $t = 0$ ).

- The blow-up  $\bar{x}$  is parabolic: where a *parabolic collision trajectory* for the cluster  $\mathbf{k}$  is the path

$$\bar{x}_i(t) = |t|^{2/(2+\alpha)} \xi_i, \quad i \in \mathbf{k}, \quad t \in \mathbb{R},$$

where  $\xi = (\xi_i)_{i \in \mathbf{k}}$  is a central configuration with  $k$  bodies.

**Proposition 8** The sequences  $x^{\lambda_n}$  and  $(dx^{\lambda_n})/(dt)$  converge to the blow-up  $\bar{x}$  and its derivative  $\dot{\bar{x}}$  respectively, in the  $H^1$ -topology. Moreover  $\bar{x}$  is a minimizing trajectory in the sense of Morse.

$$\int_0^T [\mathcal{L}(\bar{x} + \varphi) - \mathcal{L}(\bar{q})] dt \geq 0$$

for any compactly supported variation  $f$ .

### Logarithmic Type Potentials

It is not possible to define a blow-up suitable for logarithmic type potentials. Indeed, the natural scaling should be  $\bar{x}^{\lambda_n}(t) := \lambda_n^{-1} \bar{x}(\lambda_n t)$ , which does not converge, since, by the asymptotics  $|x(t)| \simeq |t| \sqrt{-\log |t|}$ , we have:

$$\lim_{\lambda_n \rightarrow 0} |\bar{x}^{\lambda_n}(t)| = \lim_{\lambda_n \rightarrow 0} \frac{r(\lambda_n t)}{\lambda_n t \sqrt{-2M(0) \log(\lambda_n t)}} \times t \sqrt{-2M(0) \log(\lambda_n t)} = +\infty$$

for every  $t > 0$ .

On the other, hand, looking at the differential equation, the (right) blow-up should be:

$$\bar{q}(t) := t\bar{s}, \quad i \in \mathbf{k},$$

where  $\bar{s}$  is a central configuration for the system limit of a sequence  $s(\lambda_n)$  where  $(\lambda_n)_n$  is such that  $\lambda_n \rightarrow 0$ . This limiting function is the pointwise limit of the normalized sequence

$$\bar{x}^{\lambda_n}(t) := \frac{1}{\lambda_n \sqrt{-2M(0) \log \lambda_n}} \bar{x}(\lambda_n t).$$

Unfortunately this path is not locally minimal for the limiting problem, indeed since, the sequence  $(\ddot{x}^{\lambda_n})_n$  converges to 0 as  $n$  tends to  $+\infty$ , and hence this blow-up minimizes only the kinetic part of the action functional.

### Absence of Collision for Locally Minimal Paths

As a matter of fact, solutions to the Newtonian  $n$ -body problem which are minimals for the action are, very likely, free of any collision. This fact was observed by the construction of suitable local variation arguments for the 2 and 3-body cases by Serra and Terracini (1992, 1994). The 4-body case was treated afterward by Dell'Antonio ((1998), with a non completely rigorous argument) and then by A. Venturelli (2002). In general, the proof goes by the sake of the

contradiction and involves the construction of a suitable variation that lowers the action in presence of a collision. A recent breakthrough in this direction is due of the neat idea, due to C. Marchal, (2002), of averaging over a family of variations parametrized on a sphere. The method of averaged variations for Newtonian potentials has been outlined and exposed by Chenciner (2002b), and then fully proved and extended to  $\alpha$ -homogeneous potentials and various constrained minimization problems by Ferrario and Terracini in (Ferrario and Terracini 2004). This technique can be used in many of the known cases to prove that minimizing trajectories are collisionless.

Of course, in some specific situations, other arguments can be useful, such as level estimates, on the infimum of the action on colliding paths (Bessi and Coti Zelati 1991; Chen 2001a, b; Chenciner 2002b). However, these argument require global conditions on the potentials and can not be applied in the present setting, where we work under local assumptions about the singularities.

We are in a position to prove the absence of collisions for locally minimal solutions when the potentials have quasi-homogeneous or logarithmic singularities. The first case is simpler, because one can take advantage the blow-up technique already exploited in Ferrario and Terracini (2004). On the other hand, when dealing with logarithmic potentials, the blow-up technique is no longer available and we conclude proving directly some averaging estimates that can be used to show the nonminimality of large classes of colliding motions. The following results are exposed in Barutello et al. (2008b).

### Quasi-Homogeneous Potentials

Let  $\tilde{U}$  be the  $C^1$  function defined on  $(a, b) \times (\mathbb{R}^k \setminus \Delta)$  in the following way:

$$\tilde{U}(t, x) = |x|^{-\alpha} \lim_{r \rightarrow 0} r^\alpha \tilde{U}(t, rx/|x|).$$

With a slight abuse of notation, we denote  $\tilde{U}(x) = \tilde{U}(r^*, x)$ .  $\tilde{U}$  is homogeneous of degree  $-\alpha$ .

[U6] there is a 2-dimensional linear subspace of  $V_{\mu(\xi)}^\perp$ , say  $W$ , where  $\tilde{U}$  is rotationally invariant:

$$\tilde{U}(e^{i\theta}w) = \tilde{U}(w), \quad \forall w \in W, \quad \forall \theta \in [0, 2\pi];$$

[U7] for every  $x \in \mathbb{R}^k$  and  $\delta \in W$  there holds:

$$\begin{aligned} & \tilde{U}(x + \delta) \\ & \leq \tilde{U} \left( \left( \frac{\tilde{U}(\pi_W(x))}{\tilde{U}(x)} \right)^{1/\alpha} \pi_W(x) + \left( \frac{\tilde{U}(x)}{\tilde{U}(\pi_W(x))} \right)^{1/\alpha} \delta \right), \end{aligned}$$

where  $\pi_W$  denotes the orthogonal projection onto  $W$ .

**Theorem 19** In addition to [U0], [U1], [U2<sub>h</sub>], [U3<sub>h</sub>], [U4<sub>h</sub>], [U5], assume that, for all  $\xi \in \Delta$  [U6] and [U7] hold. Then generalized solutions do not have collisions at the time  $t^*$ .

**Remark 2** As our potential  $\tilde{U}$  is homogeneous of degree  $-\alpha$  the function

$$\varphi(x) = \tilde{U}^{-1/\alpha}(x)$$

is a non negative, homogeneous of degree one function, having now  $\Delta$  as zero set. In most of our applications  $\varphi$  will be indeed a quadratic form. Assume that  $\varphi^2$  splits in the following way:

$$\varphi^2(x) = K|\pi_W(x)|^2 + \varphi^2(\pi_{W^\perp}(x))$$

for some positive constant  $K$ . Then [U6] and [U7] are satisfied. Indeed, denoting  $w = \pi_W(x)$  and  $z = x - w$  we have, for every  $\delta \in W$ ,

$$\begin{aligned} \varphi^2(x + \delta) &= K|w + \delta|^2 + \varphi^2(z) \\ &= K \left| \frac{\varphi(x)}{\varphi(w)}w + \frac{\varphi(w)}{\varphi(x)}\delta \right|^2 + K \frac{\varphi^2(z)}{\varphi^2(x)}|\delta|^2 \\ &\geq K \left| \frac{\varphi(x)}{\varphi(w)}w + \frac{\varphi(w)}{\varphi(x)}\delta \right|^2 \\ &= \varphi^2 \left( \frac{\varphi(x)}{\varphi(w)}w + \frac{\varphi(w)}{\varphi(x)}\delta \right), \end{aligned}$$

which is obviously equivalent to [U7].

**Proposition 9** Assume  $\tilde{U}(x) = Q^{-\alpha/2}(x)$  for some non negative quadratic form  $Q(x) = \langle Ax, x \rangle$ . Then assumptions [U6] and [U7] are satisfied whenever  $W$  is included in an Eigen space of  $A$  associated with a multiple eigenvalue.

Given two potentials satisfying [U6] and [U7] for a common subspace  $W$ , their sum enjoys the same properties. On the other hand, the class of potentials satisfying [U6] and [U7] is not stable with respect to the sum of potentials. In order to deal with a class of potentials which is closed with respect to the sum, we introduce the following variant of the last Theorem.

**Theorem 20** In addition to [U0], [U1], [U2<sub>h</sub>], [U3<sub>h</sub>], [U4<sub>h</sub>], [U5], assume that  $\tilde{U}$  has the form

$$\tilde{U}(x) = \sum_{v=1}^N \frac{K_v}{(\text{dist}(x, V_v))^\alpha},$$

where  $K_v$  are positive constants and  $V_v$  is a family of linear subspaces, with  $\text{codim}(V_v) \geq 2$ , for every  $v = 1, \dots, N$ . Then locally minimizing trajectories do not have collisions at the time  $t^*$ .

These two theorems extend to logarithmic potentials.

### Neumann Boundary Conditions and G-equivariant Minimizers

Our analysis allows to prove that minimizers to the fixed-ends (Bolza) problems are free of collisions: indeed all the variations of our class have compact support. However, other type of boundary conditions (generalized Neumann) can be treated in the same way. Indeed, consider a trajectory which is a (local) minimizer of the action among all paths satisfying the boundary conditions

$$x(0) \in X^0, \quad x(T) \in X^1,$$

where  $X^0$  and  $X^1$  are two given linear subspaces of the configuration space. Consider a (locally) minimizing path  $\bar{x}$ : of course it has not interior collisions. In order to exclude boundary collisions we

have ensure that the class of variations preserve the boundary conditions. This can be achieved by imposing assumptions [U6] and [U7] to be fulfilled also by the restriction of the potential to the boundary subspaces  $X^i$ . The analysis of boundary conditions was a key point in the paper (Ferrario and Terracini 2004), where symmetric periodic trajectories were constructed by reflections about given subspaces. Our results can be used to prove the absence of collisions also for  $G$ -equivariant (local) minimizers, provided the group  $G$  satisfies the Rotating Circle Property defined in Definition 3. Hence, existence of  $G$ -equivariant collisionless periodic solutions can be proved for the wide class of symmetry groups described in (Barutello et al. 2008a; Ferrario and Terracini 2004), for a much larger class of interacting potentials, including quasi-homogeneous and logarithmic ones. On the other hand, our results can be applied to prove that  $G$ -equivariant minimals are collisionless for many relevant symmetry groups violating the rotating circle property, such as the groups of rotations recently introduced in (Ferrario 2007, 2006).

### The Standard Variation

In order to prove that local minimizers are free of collisions we are going to make use of the following class of variations:

**Definition 7** The *standard variation* associated to  $\delta \in \Xi$  and  $T$  is defined as

$$v^\delta(t) = \begin{cases} \delta & \text{if } 0 \leq |t| \leq T - |\delta| \\ (T - t) \frac{\delta}{|\delta|} & \text{if } T - |\delta| \leq |t| \leq T \\ 0 & \text{if } |t| \geq T \end{cases}$$

Our next goal is to find a standard variation  $v^\delta$  such that the trajectory  $q + v^\delta$  does not have a collision at  $t = 0$  and

$$\Delta \mathcal{A} := \int_{-\infty}^{+\infty} [\mathcal{L}_k(q + v^\delta) - \mathcal{L}_k(q)] dt < 0$$

Let us introduce the function

$$S(\xi, \delta) = \int_0^{+\infty} \left( \frac{1}{|\xi t^{2/(2+\alpha)} - \delta|^\alpha} - \frac{1}{|\xi t^{2/(2+\alpha)}|^\alpha} \right) dt,$$

where  $\xi, \delta \in \mathbb{R}^2$ . We first estimate the variation of the action as  $\delta \rightarrow 0$ :

**Theorem 21** Let  $q = \{q\}_i = \{t^{2/(2+\alpha)}\xi_i\}$ ,  $i = 1, \dots, k$  be a parabolic collision trajectory and  $v^\delta$  a standard variation. Then, as  $\delta \rightarrow 0$

$$\begin{aligned} \Delta \mathcal{A} &= 2|\delta|^{1-\alpha/2} \sum_{\substack{i < j \\ i, j \in \mathbf{k}}} m_i m_j S\left(\xi_i - \xi_j, \frac{\delta_i - \delta_j}{|\delta|}\right) \\ &\quad + O(|\delta|). \end{aligned}$$

We observe that

$$S(\lambda\xi, \mu\delta) = |\lambda|^{-1-\alpha/2} |\mu|^{1-\alpha/2} S(\xi, \delta)$$

and hence the sign of  $S$  depends on the angle between  $\xi$  and  $\delta$ . Let

$$\begin{aligned} \Phi(\vartheta) &= \int_0^{+\infty} \frac{1}{\left(t^{\frac{4}{2+\alpha}} - 2 \cos \vartheta t^{\frac{2}{2+\alpha}} + 1\right)^{\alpha/2}} \\ &\quad - \frac{1}{t^{\frac{2\alpha}{2+\alpha}}} dt, \quad \alpha \in (0, 2). \end{aligned}$$

$\Phi(\theta)$  represents the *potential differential* needed for displacing the colliding particle from zero to  $e^{i\theta}$ . Expanding, we can express  $\Phi(\theta)$  as a hypergeometric function:

$$\begin{aligned} \Phi(\vartheta) &= \frac{\alpha(\alpha+2)}{2} \left\{ \frac{1}{\alpha-2} \beta\left(\frac{\alpha+2}{4}, \frac{\alpha+2}{4}\right) \right. \\ &\quad + \frac{1}{\alpha} \sum_{k=1}^{+\infty} \binom{-\alpha/2}{k} (-1)^k 2^{k-1} (\cos \vartheta)^k \beta \\ &\quad \left. \times \left(\frac{\alpha}{4} - \frac{1}{2} + \frac{k}{2}, \frac{\alpha}{4} + \frac{1}{2} + \frac{k}{2}\right) \right\}. \end{aligned}$$

### Some Properties of $\Phi$

The value of  $\Phi(\theta)$  ranges from  $+\infty$  to some negative value, depending on  $\alpha$ . However, thanks to

some harmonic analysis one can prove that suitable *averages* are always negative: the first inequality is particularly useful for dealing with *reflected triple collisions from the Lagrange central configuration*:

$$\Phi\left(\frac{2\pi}{3} + \gamma\right) + \Phi\left(\frac{2\pi}{3} - \gamma\right) < 0, \quad \forall \gamma \in [0, \pi/2].$$

A key remark was made by Christian Marchal: being the Newton potential a harmonic map averaging it on a sphere results in a truncation in the interior. In fact, is not so much a matter of harmonicity. A crucial estimate was proved in [FT] about the averages of  $\Phi$  on circles:

**Theorem 22** For every  $\alpha > 0$ ,  $\xi \in \mathbb{R}^3 \setminus \{0\}$  and for every circle  $\mathbb{S} \subset \mathbb{R}^d$  with center in 0,

$$\begin{aligned} \tilde{S}(\xi, \mathbb{S}) &= \frac{1}{|\mathbb{S}|} \int_{\mathbb{S}} S(\xi, \delta) d\delta \\ &= |\xi|^{-1-\alpha/2} |\delta|^{1-\alpha/2} \frac{1}{2\pi} \int_0^{2\pi} \Phi(\theta) d\theta < 0. \end{aligned}$$

Consider  $\xi = x_i - x_j$  and  $\delta$  ranging in a circle. Then the above inequality implies the following principle, a generalization of Marchal's statement in (Chenciner 2002b):

it is more convenient (from the point of view of the integral of the potential on the time line) to replace one of the point particles with a homogeneous circle of same mass and fixed radius which is moving keeping its center in the position of the original particle.

## Future Directions

This article mainly focuses on the variational approach to the search of selected trajectories to the  $n$ -body problem. In our examples, the masses can very well be equal: hence the problem can not be regarded as a small perturbation of a simpler one and a full picture of the dynamics is out of reach. In contrast, the planetary  $n$ -body problem deals with systems where one of the bodies (a “star”) is much heavier than the others (the

“planets”) and can be seen as a perturbation of a decoupled systems of one center problems. A small parameter  $\epsilon$  represents the order of the ratio between the mass of the planets and that of the star. The system is then called *nearly integrable* as it can be associated with a Hamiltonian of the form:

$$H(I, \varphi) = h(I) + \epsilon f(I, \varphi),$$

where  $I \in \mathbb{R}^N$  and  $\varphi \in \mathbb{T}^N$  ( $N$  is the number of degree of freedom) and  $\epsilon$  is a small parameter. In the three body problem we have  $N = 4$ , but the integrable limit possesses motions lying on  $\mathbb{T}^2$  (the product of two Keplerian orbits) so *the integrable limit depends on less action variables than the number of degrees of freedom*. For this reason the system is *properly degenerate*. The integrable limit possesses only quasi-periodic motions: in Jefferys and Moser (1966), the question whether these motions survive for positive values of  $\epsilon$  is settled. KAM Theory deals with the problem of persistence of such invariant tori but can not be directly applied to the planetary  $n$ -body problem, for its degeneracy. Nevertheless, the existence of invariant tori has been recently achieved in the planar and spatial three body problem (Arnold 1963; Berti et al. 2004; Biasco et al. 2003; Laskar and Robutel 1995; Robutel 1995) and in the planetary many body problem (Biasco et al. 2006; Féjoz 2004).

The future researches will extend these results to some non perturbative settings, finding both regular motions, such as periodic, quasi-periodic or quasi-periodic trajectories and irregular, chaotic ones, through the application of suitable variational methods taking into account of collision trajectories.

## Bibliography

### Primary Literature

- Albouy A (1996) The symmetric central configurations of four equal masses. *Contemp Math* 198:131–135
- Albouy A (1998) Chenciner A, Le problème des  $n$  corps et les distances mutuelles. *Invent Math* 131:151–184
- Ambrosetti A, Coti Zelati V (1993) Periodic solutions of singular Lagrangian systems. In: *Progress in nonlinear*

- differential equations and their applications, vol 10. Birkhäuser Boston Inc, Boston
- Ambrosetti A, Coti Zelati V (1994) Non-collision periodic solutions for a class of symmetric 3-body type problems. *Topol Methods Nonlinear Anal* 3(2):197–207
- Ambrosetti A, Rabinowitz PH (1973) Dual variational methods in critical point theory and applications. *J Funct Anal* 14:349–381
- Arioli G, Gazzola F, Terracini S (2000) Minimization properties of Hill's orbits and applications to some  $N$ -body problems. *Ann Inst H Poincaré Anal Non Linéaire* 17(5):617–650
- Arioli G, Barutello V, Terracini S (2006) A new branch of mountain pass solutions for the choreographical 3-body problem. *Commun Math Phys* 268(5):439–463
- Arnold VI (1963) Small denominators and problems of stability of motions in classical and celestial mechanics. *Uspehi Naut Nauk* 18(6):91–192
- Bahri A, Rabinowitz PH (1991) Periodic solutions of Hamiltonian systems of 3-body type. *Ann Inst H Poincaré Anal Non Linéaire* 8(6):561–649
- Barutello V (2004) On the  $n$ -body problem, PhD thesis, Università di Milano–Bicocca. Available at <http://www.matapp.unimib.it/dottorato/>
- Barutello V, Secchi S (2008) Morse index properties of colliding solutions to the  $n$ -body problem. *Arxiv:math/0609837*. *Ann Inst Henri Poincaré (C) Non Linear Anal* 25:539–565
- Barutello V, Terracini S (2004) Action minimizing orbits in the  $n$ -body problem with choreography constraint. *Nonlinearity* 17:2015–2039
- Barutello V, Terracini S (2007) A bisection algorithm for the numerical Mountain Pass. *NoDEA* 14:527–539
- Barutello V, Ferrario DL, Terracini S (2008a) Symmetry groups of the planar 3-body problem and action-minimizing trajectories. *Arch Ration Mech Anal* 190:189–226
- Barutello V, Ferrario DL, Terracini S (2008b) On the singularities of generalized solutions to the  $N$ -body problem. *Int Math Res Notices* 2008:rnn069–78
- Berti M, Biasco L, Valdinoci E (2004) Periodic orbits close to invariant tori and applications to the three-body problem. *Ann Scuola Norm Sup Pisa Cl Sci III(5)*:87–138
- Bessi U, Coti Zelati V (1991) Symmetries and noncollision closed orbits for planar  $N$ -body-type problems. *Nonlinear Anal* 16(6):587–598
- Biasco L, Chierchia L, Valdinoci E (2003) Elliptic two dimensional invariant tori for the planetary three-body problem. *Arch Ration Mech Anal* 170:91–135
- Biasco L, Chierchia L, Valdinoci E (2006)  $N$ -dimensional elliptic invariant tori for the planetary  $(N+1)$ -body problem. *SIAMJ Math Anal* 37:1560–1588
- Chen K-C (2001a) Action-minimizing orbits in the parallelogram four-body problem with equal masses. *Arch Ration Mech Anal* 158:293–318
- Chen K-C (2001b) On Chenciner–Montgomery's orbit in the three-body problem. *Discrete Contin Dyn Syst* 7(1):85–90
- Chen K-C (2003a) Binary decompositions for planar  $n$ -body problems and symmetric periodic solutions. *Arch Ration Mech Anal* 170(3):247–276
- Chen K-C (2003b) Variational methods on periodic and quasi-periodic solutions for the  $N$ -body problem. *Ergodic Theory Dyn Syst* 23(6):1691–1715
- Chenciner A (2002a) Action minimizing periodic orbits in the Newtonian-body problem. *Celestial Mechanics*, dedicated to Don Saari. *Contemp Math* 292:71–90
- Chenciner A (2002b) Action minimizing solutions of the Newtonian  $n$ -body problem: from homology to symmetry. *ICM, Peking*
- Chenciner A (2002c) Simple non-planar periodic solutions of the  $n$ -body problem. In: *Proceedings of the NDDS conference, Kyoto*
- Chenciner A, Desolneux N (1998) Minima de l'intégrale d'action et équilibres relatifs de  $n$  corps. *C R Acad Sci Paris Sér I* 326:1209–1212; Correction in: (1998) *C R Acad Sci Paris Sér I* 327:193
- Chenciner A, Féjoz J (2005) L'équation aux variations verticales d'un équilibre relatif comme source de nouvelles solutions périodiques du problème des  $N$  corps. *C R Math Acad Sci Paris* 340(8):593–598
- Chenciner A, Montgomery R (2000) A remarkable periodic solution of the three body problem in the case of equal masses. *Ann Math* 152(3):881–901
- Chenciner A, Venturelli A (2000) Minima de l'intégrale d'action du problème Newtonien de 4 corps de masses égales dans  $R^3$ : orbites “hip-hop”. *Celest Mech* 77:139–152
- Chenciner A, Gerver J, Montgomery R, Simó C (2001) Simple choreographies of  $N$  bodies: a preliminary study. In: *Geometry, mechanics and dynamics*. Springer, New York, pp 287–308
- Chenciner A, Féjoz J, Montgomery R (2005) Rotating eights. I. The three  $\Gamma_i$  families. *Nonlinearity* 18(3):1407–1424
- Degiovanni M, Giannoni F (1988) Dynamical systems with Newtonian type potentials. *Ann Scuola Norm Sup Pisa, Ser IV* 15:467–494
- Degiovanni M, Giannoni F, Marino A (1987) Dynamical systems with Newtonian type potentials. *Atti Accad Naz Lincei Rend Cl Sci Fis Mat Natur Ser* 8(81):271–278
- Dell'Antonio G (1998) Non-collision periodic solutions of the  $N$ -body system. *NoDEA Nonlinear Differ Equ Appl* 5(1):117–136
- Devaney RL (1978a) Collision orbits in the anisotropic Kepler problem. *Invent Math* 45(3):221–251
- Devaney RL (1978b) Nonregularizability of the anisotropic Kepler problem. *J Differ Equ* 29(2):252–268
- Devaney RL (1980) Triple collision in the planar isosceles three-body problem. *Invent Math* 60(3):249–267
- Devaney RL (1981) Singularities in classical mechanical systems. In: *Ergodic theory and dynamical systems*, I. College Park, Md., 1979–80, vol 10 of *Progr Math*. Birkhäuser, Boston, pp 211–333
- Diacu F (1992) Regularization of partial collisions in the  $N$ -body problem. *Differ Integr Equ* 5(1):103–136



- Diacu F (1993) Painlevé's conjecture. *Math Intell* 15(2): 6–12
- Diacu F (1996) Near-collision dynamics for particle systems with quasi homogeneous potentials. *J Differ Equ* 128(1):58–77
- Diacu F (2002) Singularities of the  $N$ -body problem. In: *Classical and celestial mechanics*. Princeton University Press, Princeton, pp 35–62, Recife, 1993/1999
- Diacu F, Santoprete M (2004) On the global dynamics of the anisotropic Manev problem. *Phys D* 194(1–2):75–94
- Diacu F, Pérez-Chavela E, Santoprete M (2005) The Kepler problem with anisotropic perturbations. *J Math Phys* 46(7):072701, 21
- Diacu F, Pérez-Chavela E, Santoprete M (2006) Central configurations and total collisions for quasi homogeneous  $n$ -body problems. *Nonlinear Anal* 65(7): 1425–1439
- ElBialy MS (1990) Collision singularities in celestial mechanics. *SIAM J Math Anal* 21(6):1563–1593
- Féjóz J (2004) Démonstration du “théorème d’Arnold” sur la stabilité du système planétaire (d’après Michel Herman). (french) [Proof of “Arnold’s theorem” on the stability of a planetary system (following Michel Herman)]. *Ergodic Theory Dyn Syst* 24(5):1521–1582
- Ferrario DL (2002) Symmetric periodic orbits for the  $n$ -body problem: some preliminary results. Preprint of the Max-Planck-Institut für Mathematik MPI-2002-79
- Ferrario DL (2006) Symmetry groups and non-planar collisionless action-minimizing solutions of the three-body problem in three-dimensional space. *Arch Ration Mech Anal* 179(3):389–412
- Ferrario DL (2007) Transitive decomposition of symmetry groups for the  $n$ -body problem. *Adv Math* 213: 763–784
- Ferrario D, Terracini S (2004) On the existence of collisionless equivariant minimizers for the classical  $n$ -body problem. *Invent Math* 155(2):305–362
- Gordon WB (1975) Conservative dynamical systems involving strong forces. *Trans Am Math Soc* 204: 113–135
- Gordon WB (1977) A minimizing property of Keplerian orbits. *Am J Math* 99(5):961–971
- Jefferys WH, Moser J (1966) Quasi-periodic solutions for the three-body problem. *Celest Mech Dyn Astron J* 71: 508–578
- Laskar J, Robutel P (1995) Stability of the planetary three-body problem, I: Expansion of the planetary Hamiltonian. *Celest Mech Dyn Astron* 62(3):193–217
- Levi Civita T (1918) Sur la régularisation du problème des trois corps. *Acta Math* 42:99–144
- Majer P, Terracini S (1993) Periodic solutions to some problems of  $n$ -body type. *Arch Ration Mech Anal* 124(4):381–404
- Majer P, Terracini S (1995a) On the existence of infinitely many periodic solutions to some problems of  $n$ -body type. *Commun Pure Appl Math* 48(4):449–470
- Majer P, Terracini S (1995b) Multiple periodic solutions to some  $n$ -body type problems via a collision index. *Variational methods in nonlinear analysis* (Erice, 1992). Gordon and Breach, Basel, pp 245–262
- Marchal C (2000) The family  $P_{12}$  of the three-body problem – the simplest family of periodic orbits, with twelve symmetries per period. *Celest Mech Dyn Astron* 78(1–4):279–298; (2001) New developments in the dynamics of planetary systems. *Badhofgastein, 2000*
- Marchal C (2002) How the method of minimization of action avoids singularities. *Celest Mech Dyn Astron* 83:325–353
- Mather JN, McGehee R (1974) Solutions of the collinear four body problem which become unbounded in finite time. In: *Dynamical systems, theory and applications*. Rencontres, Battelle Res Inst, Seattle, Wash., pp 573–597. *Lecture notes in Phys.*, vol 38. Springer, Berlin (1975)
- McGehee R (1974) Triple collision in the collinear three-body problem. *Invent Math* 27:191–227
- McGehee R (1986) von Zeipel’s theorem on singularities in celestial mechanics. *Expo Math* 4(4):335–345
- Moeckel R (1987) Some qualitative features of the three-body problem. In: *Hamiltonian dynamical systems*. (Boulder, 1987) vol 81 of *Contemp Math*, pp 1–22. Am Math Soc, Providence RI, 1988
- Moeckel R (1990) On central configurations. *Math Z* 205: 499–517
- Montgomery R (1998) The  $N$ -body problem, the braid group, and action-minimizing periodic solutions. *Nonlinearity* 11(2):363–376
- Montgomery R (1999) Action spectrum and collisions in the planar three-body problem. In: *Celestial mechanics*. (Evanston, 1999) vol 292 of *Contemp Math* Am Math Soc, Providence RI, 2002, pp 173–184
- Moore C (1993) Braids in classical dynamics. *Phys Rev Lett* 70(24):3675–3679
- Pacella F (1987) Central configurations and the equivariant Morse theory. *Arch Ration Mech Anal* 97:59–74
- Palais RS (1979) The principle of symmetric criticality. *Commun Math Phys* 69:19–30
- Poincaré H (1896) Sur les solutions périodiques et le principe de moindre action. *C R Acad Sci Paris, Sér I Math* 123:915–918
- Pollard H, Saari DG (1968) Singularities of the  $n$ -body problem I. *Arch Ration Mech Anal* 30:263–269
- Pollard H, Saari DG (1970) Singularities of the  $n$ -body problem II. In: *Inequalities II*. Academic, New York, pp 255–259 (Proc. Second Sympos, US Air Force Acad, Colo, 1967)
- Riahi H (1999) Study of the critical points at infinity arising from the failure of the Palais-Smale condition for  $n$ -body type problems. *Memoirs Am Math Soc* 138: 658, viii+112
- Robutel P (1995) Stability of the planetary three-body problem, II: KAM theory and existence of quasi-periodic motions. *Celest Mech Dyn Astron* 62(3): 219–261
- Saari DG (1972/73) Singularities and collisions of Newtonian gravitational systems. *Arch Ration Mech Anal* 49: 311–320

- Sbano L (1998) The topology of the planar three-body problem with zero total angular momentum and the existence of periodic orbits. *Nonlinearity* 11(3): 641–658
- Serra E, Terracini S (1992) Collisionless periodic solutions to some three-body problems. *Arch Ration Mech Anal* 120(4):305–325
- Serra E, Terracini S (1994) Noncollision solutions to some singular minimization problems with Keplerian-like potentials. *Nonlinear Anal* 22(1):45–62
- Sperling HJ (1970) On the real singularities of the  $N$ -body problem. *J Reine Angew Math* 245:15–40
- Sundman KF (1913) Mémoire sur le problème des trois corps. *Acta Math* 36:105–179
- Terracini S, Venturelli A (2007) Symmetric trajectories for the  $2N$ -body problem with equal masses. *Arch Ration Mech Anal* 184(3):465–493
- The GAP Group (2002) GAP – Groups, Algorithms, and Programming, Version 4.3, <http://www.gap-system.org>
- Venturelli A (2001) Une caractérisation variationnelle des solutions de Lagrange du problème plan des trois corps. *C R Acad Sci Paris, Sér I Math* 332(7):641–644
- Venturelli A (2002) Application de la minimisation de l'action au Problème des  $N$  corps dans le plan et dans l'espace. Thesis. University Paris VII
- Von Zeipel H (1908) Sur les singularités du problème des  $n$  corps. *Ark Math Astronom Fys* 4:1–4
- Wang Q (1991) The global solution of the  $n$ -body problem. *Celest Mech Dyn Astron* 50(1):7388
- Wintner A (1941) The analytical foundations of celestial mechanics. Princeton mathematical series, vol 5. Princeton University Press, Princeton
- Xia Z (1992) The existence of non collision singularities in Newtonian systems. *Ann Math* 135:411–468
- ## Books and Reviews
- Arnold VI, Kozlov V, Neishtadt A (2006) Mathematical aspects of classical and celestial mechanics [Dynamical systems. III], 3rd edn. In: *Encyclopaedia of mathematical sciences*, vol 3. Springer, Berlin, xiv and p 518, translated from the Russian original by Khukhro E
- Diacu F, Holmes P (1996) *Celestial encounters. The origins of chaos and stability*. Princeton University Press, Princeton
- Meyer KR (1999) *Periodic solutions of the  $N$ -body problem*. Lecture notes in mathematics, 1719. Springer, Berlin
- Moser J (1973) *Stable and random motions in dynamical systems*, With special emphasis on celestial mechanics, Hermann Weyl Lectures, the Institute for Advanced Study, Princeton NJ, *Annals of Mathematics Studies*, No. 77. Princeton University Press, Princeton
- Pollard H (1976) *Celestial mechanics*, Carus mathematical monographs, 18. Mathematical Association of America, Washington, DC
- Saari D (2005) *Collisions, rings, and other Newtonian  $N$ -body problems*, CBMS regional conference series in mathematics, 104. American Mathematical Society, Providence RI, Washington, Published for the Conference Board of the Mathematical Sciences
- Siegel CL, Moser JK (1995) *Lectures on celestial mechanics*. In: *Classics in mathematics*. Springer, Berlin, Translated from the German by Kalme CI, Reprint of the 1971 translation
- Stiefel EL, Scheifele G (1971) *Linear and regular celestial mechanics. Perturbed two-body motion, numerical methods, canonical theory*. Die Grundlehren der mathematischen Wissenschaften, Band 174. Springer, New York/Heidelberg

# Semiclassical Perturbation Theory

Andrea Sacchetti

Dipartimento di Matematica Pura ed Applicata,  
Università di Modena e Reggio Emilia, Modena,  
Italy

## Article Outline

Glossary

Definition of the Subject

Introduction

The WKB Approximation

Semiclassical Approximation in Any Dimension

Propagation of Quantum Observables

Future Directions

Bibliography

## Glossary

**Agmon metric** In the classically forbidden region where the potential energy  $V(x)$  is larger than the total energy  $E$ , i.e.  $V(x) > E$ , we introduce a notion of distance based on the *Agmon metric* defined as  $[V(x) - E]dx^2$ , where  $dx^2$  is the usual Riemann metric. We emphasize that such an Agmon metric is the “semiclassical” equivalent of the “classical” Jacobi metric  $[E - V(x)]dx^2$  introduced in classical mechanics in the classically permitted region where  $V(x) < E$ .

**Asymptotic series** The notion of asymptotic series goes back to Poincaré. We say that a formal power series  $\sum_r a_r z^{-r}$  is taken to be the asymptotic power series for a function  $f(z)$ , as  $|z| \rightarrow \infty$  in a given infinite sector  $S = \{z \in \mathbb{C} : \alpha \leq \arg z \leq \beta\}$ , if for any fixed  $N$  the remainder term

$$R_N(z) = f(z) - \sum_{r=0}^{N-1} a_r z^{-r}$$

is such that  $R_N(z) = O(z^{-N})$ , that is

$$|R_N(z)| \leq C_N |z|^{-N}, \forall z \in S,$$

for some positive constant  $C_N$  depending on  $N$ .

**Classically allowed and forbidden regions – turning points** We distinguish two different regions: the region  $V(x) < E$  where the classical motion is allowed and the region  $V(x) > E$  where the classical motion is forbidden. The points that separate these two regions, that is such that  $V(x) = E$ , will play a special role and they are named turning points.

**Semiclassical limit** Cornerstone of Quantum Mechanics is the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V(x)\psi, \quad (1)$$

where  $V(x)$  is assumed to be a real-valued function, usually it represents the potential energy, and  $m$  is the mass of the particle. The solution of this equation defines the density of probability  $|\psi(x, t)|^2$  to find the particle in some region of space. The parameter  $\hbar$  in Eq. (1) is related to Planck’s constant  $h$

$$\begin{aligned} \hbar &= \frac{h}{2\pi} = 1.0545 \times 10^{-27} \text{ erg sec} \\ &= 6.5819 \times 10^{-22} \text{ MeV sec.} \end{aligned}$$

According to the correspondence principle, when Planck’s constant can be considered small with respect to the other parameters, such as masses and distances, then quantum theory approaches classical Newton theory. Thus, roughly speaking, we expect that classical mechanics is contained in quantum mechanics as a limiting form (i.e.  $\hbar \rightarrow 0$ ).

The limit of small  $\hbar$ , when compared with the other parameters, is the so-called semiclassical limit. We should emphasize that making Planck's constant small in Eq. (1) is a rather singular limit and many difficult mathematical problems occur.

**Stokes lines** There is some misunderstanding in the literature concerning the name of the curves  $\mathcal{I}[\xi(z)] = 0$ , where  $\mathcal{I}(\xi)$  denotes the imaginary part of

$$\xi(z) = \frac{i}{\hbar} \int_{x_E}^z \sqrt{E - V(q)} dq, \quad z \in \mathbb{C},$$

and  $x_E \in \mathbb{R}$  is a turning point (i.e.  $V(x_E) = E$ ), the potential  $V$  is assumed to be an analytic function in the complex plane. As usually physicists do, and in agreement with Stokes' original treatment, we adopt here the convention to name as Stokes line any path coming from the turning point  $x_E$  such that  $\mathcal{I}[\xi(z)] = 0$ ; reserving the name of anti-Stokes lines, or regular lines, for the curves such that  $\Re[\xi(z)] = 0$ , where  $\Re(\xi)$  denotes the real part of  $\xi$  (we should emphasize that most mathematicians adopt the opposite rule, that is they call Stokes lines the paths such that  $\Re[\xi(z)] = 0$ ).

**Tunnel effect** At a real (simple) turning point  $x_E$  where  $E = V(x)$  (and  $V'(x) \neq 0$ ), classical particles incident from an accessible region [where  $E > V(x)$ ] reverse their velocity and return back; the adjacent region [where  $E < V(x)$ ] is forbidden to these particles according to classical mechanics. In fact, quantum mechanically exponentially growing or damped waves can exist in forbidden regions and a quantum particle can pass through a potential barrier. This is the so-called tunnel effect.

**WKB** The WKB method, named after the contributions independently given by Wentzel, Kramers and Brillouin, consists of connecting the approximate solutions of the time independent Schrödinger equation across turning points.

## Definition of the Subject

Several kinds of perturbation methods are commonly used in quantum mechanics (► [“Perturbation Theory and Molecular Dynamics”](#), ► [“Perturbation Theory in Quantum Mechanics”](#), ► [“Perturbation Theory”](#)). This entry deals with *semiclassical approximations* where expressions for energy levels and wave functions are obtained in the limiting case of small values of Planck's constant.

We emphasize that wave functions are highly singular as the parameter  $\hbar$  goes to zero. In fact the semiclassical limit is a singular perturbation problem; namely, Eq. (1) suffers a reduction of order setting  $\hbar$  equal to zero and the resulting equation is not a differential equation and does not give the classical limit correctly. Therefore, ordinary perturbation methods, which usually give energy levels and wave-functions as convergent power series of the small parameter, cannot be applied.

The main goal of semiclassical methods consists of obtaining *asymptotic series*, in the limit of small  $\hbar$ , for the quantum-mechanical quantities. For instance, semiclassical methods permit us to solve the eigenvalues problem

$$-\frac{\hbar^2}{2m} \Delta \psi + V(x) \psi = E \psi, \quad x \in \mathbb{R}^n, \quad n \geq 1, \quad (2)$$

where the energy levels  $E$  and the wave-functions  $\psi(x)$  are given by means of asymptotic series in the limit of vanishing  $\hbar$ , or to obtain a direct link between the time evolution of a quantum observable and the Hamiltonian flux of the associated classical quantity.

## Introduction

The first contributions to this theory, in the framework of Quantum Mechanics, go back to Wentzel (1926), Kramers (1926) and Brillouin (1926). In their papers they independently developed the determination of connection formulas linking exponential and oscillatory approximations of the solution of Eq. (2) in dimension  $n = 1$  across a turning point. This method, explained in section

**The WKB Approximation**, is usually called WKB approximation, or also JWKB method to acknowledge that the approximate connection formula was previously discovered by Jeffrey. Actually, oscillatory and exponential approximation formulas, obtained far from turning points, were independently used by Green (1837) and Liouville (1837) and they can be already found in an investigation on the motion of a planet in an unperturbed elliptic orbit by the Italian astronomer Carlini (1817). For historical notes on the WKB-method we refer to Dingle (1973), Fröman and Fröman (2002), and McHugh (1971).

The WKB method can be also applied to three-dimensional problems only under some particular circumstances; for instance when the potential is spherically symmetric and the radial differential equation can be separated.

In general WKB approximation is not suitable for problems in dimension  $n$  higher than 1. Actually, semiclassical methods in higher dimension require new sophisticated tools such as the *Agmon metric*, *microlocal calculus* and  *$\hbar$ -pseudo-differential operators*. These tools have been developed by Agmon, Hörmander and Maslov in the 1970s and since then a large number of mathematicians have contributed to this subject. In sections “[Semiclassical Approximation in Any Dimension](#)” and “[Propagation of Quantum Observables](#)” we briefly introduce the reader to the basic concepts of these theories and we resume the most important results such as the exponential decay of wave-functions and the Egorov Theorem, which asymptotically describes the quantum evolution of an observable by means of the classical evolution of its classical counterpart.

Actually, these methods may be applied to many other fields, where  $\hbar$  is a small quantity not related to the Planck constant but it may represent a different small physical quantity such as the adiabatic parameter in adiabatic problems, the (square root of the) inverse of the heavy mass in the Born–Oppenheimer approximation, etc.

## Notation

Hereafter, for the sake of definiteness, let us fix  $2m = 1$ .

## The WKB Approximation

If the potential  $V(x)$  does not have a very simple form then the solution of the time-independent Schrödinger equation even in one dimension

$$-\hbar^2 \psi'' + V(x)\psi = E\psi, \quad x \in (x_1, x_2), \quad (3)$$

is a quite complicated problem which requires the use of a sort of approximation method (here after  $' = d/dx$  denotes the derivative with respect to  $x$ ); here  $(x_1, x_2)$  is a given finite or infinite one-dimensional interval.

We distinguish different regions: the classically allowed regions where  $V(x) < E$  and the classically forbidden regions where  $V(x) > E$ . Approximation formulas for the solution of Eq. (3) in these separate regions have been studied since Carlini, Green, Liouville and Jacobi. The problem to connect these approximated solutions across the turning points was raised in the framework of Quantum Mechanics and it was independently solved by Jensen, Wentzel, Kramers and Brillouin. For the general treatment of semiclassical approximation in dimension one and for physical applications we refer to Berezin and Shubin (1991), Berry and Mount (1972), Fröman and Fröman (1965), Landau and Lifshitz (1959), Merzbacher (1970), Olver (1974), and Voros (1982).

## Semiclassical Solutions in the Classically Allowed Region

The basic idea is quite simple: if  $V = \text{const.}$  is a constant smaller than  $E$  then Eq. (3) has solutions  $e^{\pm ikx/\hbar}$  for a suitable real-valued constant  $k$ . This fact suggests to us that if the potential, while no longer constant, varies only slowly with  $x$  and it is such that  $V(x) < E$  for any  $x \in (x_1, x_2)$ , we might try a solution of the form

$$\psi(x) = a(x)e^{iS(x)/\hbar}, \quad (4)$$

except that the *amplitude*  $a(x)$  is not constant and the *phase*  $S(x)$  is not simply proportional to  $x$ . Substituting Eq. (4) into Eq. (3) and separating the real and imaginary parts of the coefficients of  $e^{iS(x)/\hbar}$ , then we get a system of equations for the

$x$ -dependent real-valued phase  $S(x)$  and amplitude  $a(x)$ :

$$\begin{cases} u^2 - p^2(x) = \hbar^2 a''/a \\ (a^2 u)' = 0 \end{cases}$$

where we set

$$u(x) = S'(x) \quad \text{and} \quad p(x) = \sqrt{E - V(x)}.$$

Hence,  $a(x) = C[u(x)]^{-1/2}$  for some constant  $C$ .

Since the complex conjugation  $a(x)e^{-iS(x)/\hbar}$  of Eq. (4) is a solution of the same equation and the Wronskian between these two solutions is not zero then the general solution of Eq. (3) can be written as

$$\begin{aligned} \psi(x) &= b_+ \psi_+(x) + b_- \psi_-(x), \\ \psi_{\pm}(x) &= \frac{1}{\sqrt{u(x)}} e^{\pm i/\hbar \int u(x) dx}. \end{aligned} \quad (5)$$

Here,  $b_{\pm}$  are arbitrary constants and  $u = u(x; \hbar)$  is a real-valued solution, depending on the variable  $x$  and on the semiclassical parameter  $\hbar$ , of the nonlinear ordinary differential equation

$$\mathcal{F}(u) = u^2 - p^2 - \hbar^2 u f(u) = 0 \quad (6)$$

where

$$f(u) = u^{-1/2} \left( u^{-1/2} \right)'.$$

The fact that Eq. (6) is a nonlinear equation, whereas the Schrödinger equation (3) is linear, would be usually regarded as a drawback, but we shall take advantage of the nonlinearity to develop a simple approximation method for solving Eq. (6).

Taking the limit of  $\hbar$  small in Eq. (6) then it turns out that the leading order of  $u$  is simply given by  $p$ . Actually, it is possible to prove that Eq. (3) has twice continuously differentiable solutions  $\psi_{\pm}$  satisfying the following asymptotic behavior

$$\begin{aligned} \psi_{\pm}(x) &= \frac{1}{\sqrt{p(x)}} e^{\pm i/\hbar \int_a^x p(q) dq} [1 + \delta_{\pm}(x)], \\ x &\in (x_1, x_2), \end{aligned} \quad (7)$$

where  $a$  is an arbitrary point in  $(x_1, x_2)$  and where

$$|\delta_{\pm}(x)| \leq e^{\frac{1}{2}\hbar \left| \int_a^x |f[p(q)]| dq \right|} - 1 = O(\hbar).$$

Henceforth, the dominant term of the solution (5) has an oscillating behavior of the form

$$\begin{aligned} \psi(x) &= b_+ \frac{e^{i/\hbar \int_a^x \sqrt{E-V(q)} dq}}{\sqrt[4]{E-V(x)}} \\ &+ b_- \frac{e^{-i/\hbar \int_a^x \sqrt{E-V(q)} dq}}{\sqrt[4]{E-V(x)}} + O(\hbar). \end{aligned}$$

In order to compute also the other terms of the asymptotic expansion of the solutions  $\psi_{\pm}(x)$  we formally solve the nonlinear Eq. (6) by means of the formal power series in  $\hbar^2$ :

$$u = u(x, \hbar) = \sum_{n=0}^{\infty} \hbar^{2n} u_{2n}(x). \quad (8)$$

The functions  $u_{2n}(x)$  are determined explicitly, order by order, by formally substituting Eq. (8) into Eq. (6) and requiring that the coefficients of the same terms  $\hbar^{2n}$  are zero for any  $n \geq 0$ . In such a way and assuming that the potential  $V(x)$  admits derivatives at any order then we obtain that

$$\begin{aligned} u_0(x) &= p(x) \\ u_2(x) &= -\frac{1}{4} \frac{p''(x)}{p^2(x)} + \frac{3}{8} \frac{[p'(x)]^2}{p^3(x)} \\ u_4(x) &= \frac{1}{16} \frac{p'''(x)}{p^4(x)} - \frac{5}{8} \frac{p'''(x)p'(x)}{p^5(x)} - \frac{13}{32} \frac{[p''(x)]^2}{p^5(x)} \\ &\quad + \frac{99}{32} \frac{p''(x)[p'(x)]^2}{p^6(x)} - \frac{297}{128} \frac{[p'(x)]^4}{p^7(x)}, \end{aligned}$$

and so on. Thus, the asymptotic behavior (7) can be improved up to any order. In particular, let



$$\psi_{N,\pm}(x) = \frac{1}{\sqrt{p_N(x)}} e^{\pm i/\hbar \int_a^x p_N(q) dq}, \quad x \in (x_1, x_2),$$

where  $a \in (x_1, x_2)$  is arbitrary and fixed and  $p_N(x) = \sum_{n=0}^N u_{2n}(x) \hbar^{2n}$ , in particular  $p_0(x) = p(x)$ , let us introduce the *error-control function*

$$\epsilon_N(x) = \frac{1}{\hbar p_N(x)} \mathcal{F}[p_N(x)] = O(\hbar^{2N+1})$$

where  $\epsilon_0(x) = \hbar f[p(x)]$ ; then the two functions  $\psi_{N,\pm}(x)$  approximate the solutions  $\psi_{\pm}(x)$  of Eq. (3) up to the order  $2N + 1$ , that is

$$|\psi_{\pm}(x) - \psi_{N,\pm}(x)| \leq \left[ \exp \left( \frac{1}{2} \left| \int_a^x \epsilon_N(q) |dq| \right| \right) - 1 \right] = O(\hbar^{2N+1}), \quad \forall x \in (x_1, x_2).$$

We underline that this result is valid whether or not  $x_1$  and  $x_2$  are finite, also whether or not  $V(x)$  is bounded: it suffices that the *error-control function* is absolutely integrable:  $\epsilon_N \in L^1(x_1, x_2)$ .

### Semiclassical Solutions in the Classically Forbidden Region

The previous asymptotic computation of the solutions  $\psi_{\pm}$  applies also in the classically forbidden region where  $V(x) > E$ , provided that  $p(x)$  is one of the two purely imaginary determination of  $\sqrt{E - V(x)}$ . By assuming that  $\mathcal{I}p(x) < 0$  then Eq. (3) has twice continuously differentiable solutions of the form

$$\psi_{\pm}(x) = \frac{1}{\sqrt{|p(x)|}} e^{\pm 1/\hbar \int_a^x |p(q)| dq} [1 + \delta_{\pm}(x)], \quad x \in (x_1, x_2), \quad (9)$$

where  $a$  is an arbitrary point in  $(x_1, x_2)$  and where the remainder terms  $\delta_{\pm}$  are bounded as follows

$$|\delta_{\pm}(x)| \leq e^{\frac{1}{2}\hbar \left| \int_a^x |f[p(q)]| dq \right|} - 1 = O(\hbar)$$

where  $a_+ = x_1$  and  $a_- = x_2$ . In particular, if  $V(x) > E$  for any  $x \in (-\infty, x_2)$ , that is  $x_1 = -\infty$  (and similarly we can consider the case where  $x_2 = +\infty$ ), and the error-control function  $f[p(x)] \in L^1(-\infty, x_2)$  then the above asymptotic estimate holds for any  $x \in (-\infty, x_2)$ . We also emphasize that the asymptotic behavior (9) could be also improved up to any order, as done for the approximated solutions in the classically allowed regions.

We underline that the two solutions  $\psi_{\pm}$  play here a different role. Indeed, solution  $\psi_{-}(x)$  is an exponentially decreasing function as  $x > a$  grows, while  $\psi_{+}(x)$  is exponentially increasing. The first solution is usually called *recessive* solution and it is *uniquely determined* by the asymptotic power expansion as  $\hbar \rightarrow 0$ . The other solution is called *dominant* solution and it is *not uniquely determined* by the asymptotic expansion; in fact,  $\psi_{+}(x) + c\psi_{-}(x)$ ,  $x > a$ , is still a solution of Eq. (3) which has the same asymptotic behavior of  $\psi_{+}(x)$  for any  $c \in \mathbb{C}$ .

### Connection Formula

It is not possible to use the semiclassical solutions (7) and (9) at turning points  $x_E$  since the term  $1/p$  diverges and there is no guarantee that the same combination of simple semiclassical solutions will fit the same particular solution on both sides of the turning point. This is the so-called *connection problem*: if the values  $b_{\pm}$  of the semiclassical solution are known in a given region then the problem consists of computing the values of  $b_{\pm}$  in a different region separated from the first region by one (or more) turning points.

The main two methods used to treat this problem are the following ones:

- The *complex method*, where the two regions surrounding the turning point are joined by a path in the complex plane which is sufficiently far from the turning points. In such a case the semiclassical solutions in different regions are connected by means of holomorphic extension arguments. Thus, in order to apply this method we have to assume that the potential  $V(x)$  can be holomorphically extended to the complex plane.
- The second method employs the technique of the *uniform approximation*, where the required

solution is mapped on the solution of a simpler and suitable equation which has the same disposition of turning points as the original one. As well as enabling the connection problem to be solved, this method also provides the wave function in the neighborhood of the turning points, which was bypassed in the complex method.

We are going now to explain these methods in detail.

### The Complex Method

By assuming that  $V(x)$  is the restriction on the real axis of a holomorphic function  $V(z)$ ,  $z \in \mathbb{C}$ , we turn now to the approximate solution of the differential equation

$$-\hbar^2 \frac{d^2 \psi(z)}{dz^2} = [E - V(z)] \psi(z)$$

in a complex domain  $\mathcal{D}$  in which  $V(z)$  is holomorphic and  $p^2(z) = E - V(z)$  does not vanish. Then, this equation has two holomorphic solutions

$$\psi_{\pm}(z) = \frac{1}{\sqrt{p(z)}} e^{\pm i/\hbar \int_a^z p(q) dq} [1 + o(\hbar)], z \in \mathcal{D}, \quad (10)$$

where  $a$  is an arbitrary point; in particular, for our purposes it is convenient to choose it coinciding with the turning point, i.e.:  $a = x_E$ . Actually, this asymptotic behavior holds under a technical assumption: given a reference point  $a_+$  (respectively  $a_-$ ) then Eq. (10) is true for  $\psi_+(z)$  (resp.  $\psi_-(z)$ ) for any  $z$  such that there exists a *progressive* (resp. *regressive*) path  $\Gamma_+$  (resp.  $\Gamma_-$ ) contained in  $\mathcal{D}$  and connecting  $z$  with  $a_+$  (resp.  $a_-$ ); that is as  $q$  passes along  $\Gamma_+$  (resp.  $\Gamma_-$ ) from  $a_+$  (resp.  $a_-$ ) to  $z$  then  $\Re[\xi(z)]$  is nondecreasing (resp. nonincreasing) where

$$\xi(z) = \frac{i}{\hbar} \int_{x_E}^z p(q) dq.$$

Now, we denote as a Stokes line any path coming from the turning point  $x_E$  such that  $\Im[\xi(z)] = 0$ ; we denote as an anti-Stokes line, or *principal line*,

any path coming from the turning point  $x_E$  such that  $\Re[\xi(z)] = 0$ . Classically forbidden (resp. allowed) regions are particular cases of Stokes (resp. anti-Stokes) lines, which lie on the real axis.

Let us consider now the case, as in Fig. 1, where the turning point  $x_E$  is simple (that is  $V'(x_E) \neq 0$ ),  $p^2(x) < 0$  for  $x < x_E$  and  $p^2(x) > 0$  for  $x > x_E$ . Since the turning point  $x_E$  is simple then in a neighborhood of the turning point we have that  $p(z) \approx c[z - x_E]^{\frac{1}{2}}$ , for some  $c \in \mathbb{R}^+$ , and  $\xi(z) \approx \frac{2ic}{3\hbar} [z - x_E]^{\frac{3}{2}}$ . Thus, Stokes lines are three different paths coming from the turning point and with asymptotic directions  $\arg(z - x_E) = \frac{1}{3}\pi, \pi, \frac{5}{3}\pi$ ; while the asymptotic directions of the anti-Stokes lines are given by  $\arg(z - x_E) = 0, \frac{2}{3}\pi, \frac{4}{3}\pi$ .

Along the three anti-Stokes lines  $\gamma_1, \gamma_2$  and  $\gamma_3$  the solution  $\psi$  can be written as

$$\psi(z) = b_{+,j} \psi_{+,j}(z) + b_{-,j} \psi_{-,j}(z), z \in \gamma_j, j = 1, 2, 3,$$

where  $\psi_{\pm,j}$  have the asymptotic behavior given by Eq. (10) and where the coefficients  $b_{\pm,j}$  are asymptotically fixed along the anti-Stokes lines. Let  $F_j$  be the  $2 \times 2$  matrix which connects the coefficients:

$$\begin{pmatrix} b_{+,j+1} \\ b_{-,j+1} \end{pmatrix} = F_j \begin{pmatrix} b_{+,j} \\ b_{-,j} \end{pmatrix}.$$

In the region I, with boundaries given by the anti-Stokes lines  $\gamma_1$  and  $\gamma_2$ , the solution  $\psi_+$  is *dominant* and thus the coefficient  $b_+$  cannot change in this region, similarly in the region III with boundaries  $\gamma_3$  and the cut. In contrast, in the region II, with boundaries given by the anti-Stokes lines  $\gamma_2$  and  $\gamma_3$ , the solution  $\psi_-$  is *dominant* and thus the coefficient  $b_-$  cannot change in this region. Then the matrices  $F_j$  can be written as

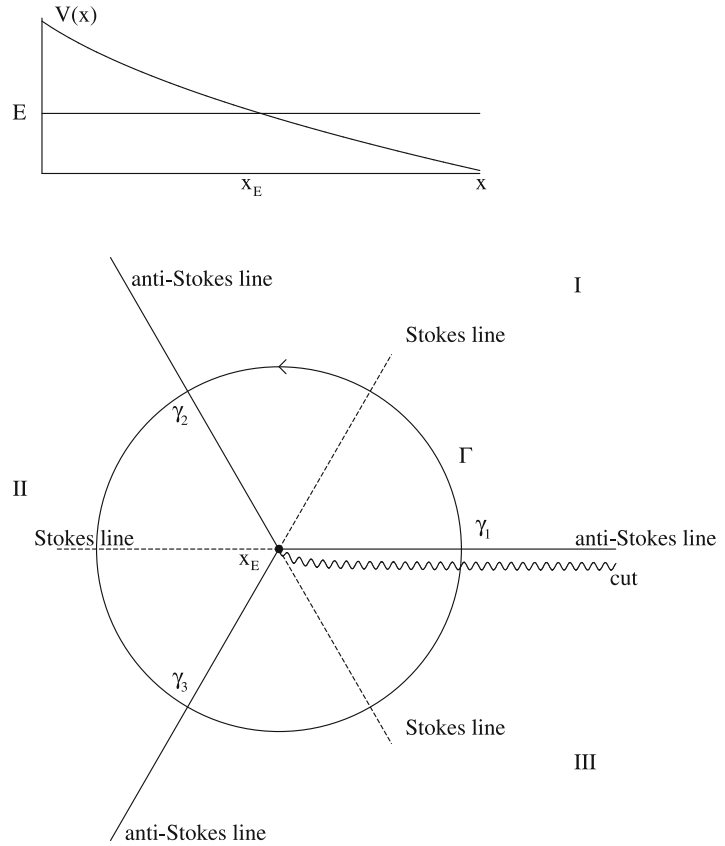
$$F_1 = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$

and  $F_3 = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$

where  $r, s$  and  $t$  have to be determined. To this end let us consider the closed anti-clockwise path  $\Gamma$  surrounding the simple turning point  $x_E$ ; along

**Semiclassical****Perturbation Theory,**

**Fig. 1** Stokes and anti-Stokes lines in a neighborhood of a simple turning point  $x_E$  where the left-hand side of the turning point is classically forbidden and the right-hand side is classically allowed. The wavy line denotes the *cut* of the multi-valued function  $p(z)$



this path the exponential term gains a phase  $v = \frac{1}{\hbar} \oint_{\Gamma} p(z) dz$  and the argument of  $P(z)$ , which appears in Eq. (10), decreases of  $\pi$ . Hence, the holomorphic extension of  $\psi_{\pm}$ , around this closed path, are thus proportional to the solution  $\psi_{\mp}$ :

$$\psi_{+} \rightarrow i e^{iv} \psi_{-} \quad \text{and} \quad \psi_{-} \rightarrow i e^{-iv} \psi_{+}.$$

Therefore, the numbers  $r, s$  and  $t$  are such that

$$\begin{aligned} F_1 F_2 F_3 &= \begin{pmatrix} 1 + st & s \\ r + (rs + 1)t & rs + 1 \end{pmatrix} \\ &\sim \begin{pmatrix} 0 & i e^{-iv} \\ i e^{iv} & 0 \end{pmatrix} \end{aligned}$$

from which it follows that

$$s = i e^{-iv}, \quad r = t = i e^{iv}.$$

In particular, since  $\int p(z) dz$  does not diverge at  $x_E$  then we can take  $\Gamma$  in a small neighborhood of

$x_E$  and  $v \rightarrow 0$ . Thus, the Stokes' rule follows: *the connection between the two coefficients from one Stokes line to the (anti-clockwise) adjacent one follows the following rule*

$$\begin{aligned} b_{\text{dominant}} &\rightarrow b_{\text{dominant}} \quad \text{and} \\ b_{\text{recessive}} &\rightarrow b_{\text{recessive}} + i b_{\text{dominant}}. \end{aligned}$$

**The Method of Comparison Equations**

With this method we obtain a local approximate solution, in a neighborhood of the turning points, in terms of the known solutions of an equivalent equation

$$\frac{d^2 \phi}{dy^2} + q(y) \phi(y) = 0, \quad (11)$$

where  $q(y)$  is chosen to be similar in some way to  $p^2(x)$ , but simpler in order to have an explicit solution. For the sake of definiteness, we restrict here our analysis to the case of a simple turning point  $x_E$ .

In order to get an approximate solution to Eq. (3) in a small neighborhood of a simple zero  $x_E$  of  $p^2(x)$  we introduce the following changes of variable

$$x \rightarrow y(x) = \left[ \frac{3}{2} \zeta(x) \right]^{\frac{2}{3}}, \quad \zeta(x) = \frac{i}{\hbar} \int_{x_E}^x p(q) dq.$$

Equation (3) takes the form of the approximate Eq. (11) where  $q(y) \approx -y$  in a neighborhood of the turning point, which admits solutions given by the Airy functions  $Ai(y)$  and  $Bi(y)$ . Then, the approximate solution of the Schrödinger equation (3) in a neighborhood of the turning point has the form

$$\psi(x) \sim \alpha \left[ \frac{\pi^2 y(x)}{p^2(x)} \right]^{\frac{1}{4}} \{ \cos(\mu) Ai[y(x)] + \sin(\mu) Bi[y(x)] \}$$

where  $\alpha$  and  $\beta$  are constants. The Airy functions are well understood and exact connection formulas can be established obtaining that

$$\psi(x) \rightarrow \frac{\alpha}{2\sqrt{|p(x)|}} \left[ \cos(\mu) e^{-|\xi|} + 2 \sin(\mu) e^{|\xi|} \right]$$

from the turning point to the classically forbidden region, and

$$\psi(x) \rightarrow \frac{\alpha}{\sqrt{|p(x)|}} \cos \left( |\xi| + \mu - \frac{1}{4}\pi \right)$$

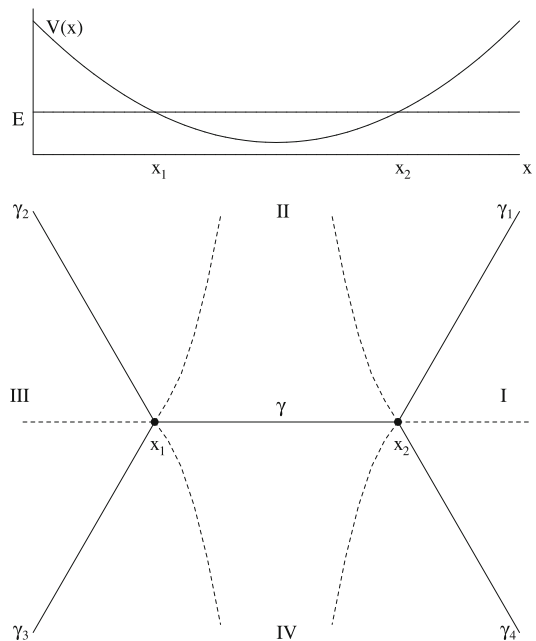
from the turning points to the classically allowed region. Thus, we have established a connection between oscillatory and exponentially increasing and decreasing solutions. Clearly, this entire approach breaks down if the energy is too close to a value corresponding to a stationary point of the potential. In this case a different approximation must be implemented. For instance, in the case of a turning point with multiplicity 2 then the approximate solution of the Schrödinger equation (3) is given by means of Weber parabolic cylinder functions.

### Bound States for a Single Well Potential

We apply now the previous techniques in order to compute the stable states, that is normalized solutions of the Eq. (2), when the potential  $V(x)$  has a single well shape; that is it has a simple minimum point  $x_{\min}$ , with minimum value  $V_{\min}$  such that  $V(x) > V_{\min}$  for any  $x \neq x_{\min}$  and  $\liminf_{x \rightarrow +\infty} V(x) > V_{\min}$ . Then, for  $E > V_{\min}$  close enough to the bottom of the well, equation  $p(x) = 0$  has only two simple solutions  $x_1 < x_{\min} < x_2$  and the typical picture of the Stokes lines appears as in Fig. 2. Here,  $\psi_{\pm, x_1}$  (resp.  $\psi_{\pm, x_2}$ ) denotes the fundamental solutions with asymptotic behaviors (7) and (9) in a neighborhood of  $x_1$  (resp.  $x_2$ ). Since we are looking for normalized solutions then in the classically forbidden region  $x > x_2$ , contained in the region I, the *dominant* term  $\psi_{+, x_2}$  of the solution  $\psi$  should have coefficient  $b_{+, x_2}^I$  exactly zero, that is

$$\psi(z) = \psi_{-, x_2}(z), \quad z \in I,$$

where the coefficients  $b_{-, x_2}$  of the *recessive* solution  $\psi_{-, x_2}$  is chosen equal to 1. Turning around the



**Semiclassical Perturbation Theory, Fig. 2** Stokes and anti-Stokes lines in a neighborhood of the bottom of a single well

turning point  $x_2$  and crossing the anti-Stokes line it follows that the coefficients of the solution  $\psi = b_{+,x_2}^{\text{II}} \psi_{+,x_2}(x) + b_{-,x_2}^{\text{II}} \psi_{-,x_2}(x)$  in the region II take the form

$$\begin{pmatrix} b_{+,x_2}^{\text{II}} \\ b_{-,x_2}^{\text{II}} \end{pmatrix} = F_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

where

$$F_2 \sim \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}$$

according to the Stokes' rule.

In order to study the matching condition around the other turning point  $x_1$  we have to transport the origin of integration at  $x_1$  obtaining

$$\psi(z) = b_{+,x_1}^{\text{II}} \psi_{+,x_1}(z) + b_{-,x_1}^{\text{II}} \psi_{-,x_1}(z)$$

where

$$\begin{aligned} b_{+,x_1}^{\text{II}} &= \psi_{+,x_2}(x_1) b_{+,x_2}^{\text{II}} \quad \text{and} \quad b_{-,x_1}^{\text{II}} \\ &= \psi_{-,x_2}(x_1) b_{-,x_2}^{\text{II}}. \end{aligned}$$

On the other hand, also in region III the *dominant* term  $\psi_{+,x_1}$  of the solution  $\psi$  should have coefficient  $b_{+,x_1}^{\text{III}}$  exactly zero; from this fact and from the Stokes' rule applied to the turning point  $x_1$  it follows that the equation for the dominant term of the bound states is

$$\psi_{+,x_2}(x_1) + \psi_{-,x_2}(x_1) = 0,$$

which implies

$$\cos \left[ \frac{1}{\hbar} \left( \int_{x_1}^{x_2} p(x) dx + O(\hbar) \right) \right] = 0.$$

Hence, we obtain the well known Bohr-Sommerfeld rule

$$\int_{x_1}^{x_2} p(x) dx = \hbar \left[ \frac{1}{2} \pi + n\pi \right] + O(\hbar^2), \quad n \in \mathbb{N}.$$

### Double Well Model: Estimate of the Splitting and the "Flea of the Elephant"

We consider now the case of a symmetric double well potential; that is  $V(-x) = V(x)$  and it has two simple minima at  $x_+ > 0$  and  $x_- = -x_+$  separated by a barrier, we assume also that the minimum value  $V_{\min} < \liminf_{|x| \rightarrow \infty} V(x)$ . For instance,  $V(x) = x^4 - 2\alpha x^2$  for some  $\alpha > 0$ ; in this case the shape of the potential has two symmetric wells where  $x_{\pm} = \pm\sqrt{\alpha}$  and  $V_{\min} = -\alpha^2 < 0$ , the two wells are separated by an energy barrier with top  $V_{\max} = 0$  at  $x = 0$ . The semiclassical double well model is not only a very enlightening pedagogical problem, but it is also the basic argument explaining many relevant physical questions, see, e.g., Claviere and Jona Lasinio (1986), Graffi et al. (1984), Harrell (1980), Simon (1985), and Wilkinson and Hannay (1987).

In the interval  $(V_{\min}, V_M)$ , where

$$V_M = \min \left[ V_{\max}, \liminf_{|x| \rightarrow \infty} V(x) \right],$$

the stable states appear as *doublets*  $E_{\pm}$  whose distance  $\omega$ , named *splitting*, is quite small. The associated eigenvectors are even and odd (real-valued) wave-functions, that is

$$\psi_{\pm}(-x) = \pm \psi_{\pm}(x). \quad (12)$$

The splitting  $\omega = E_- - E_+$  can be computed as

$$\omega = \hbar^2 \frac{\psi_+(0) \psi'_-(0)}{\int_0^{\infty} \psi_-(x) \psi_+(x) dx}$$

and it turns out to be exponentially small as  $\hbar \rightarrow 0$ . More precisely, if  $E_+$  is the ground state then  $\omega = O(e^{-S_0/\hbar})$  where

$$S_0 = \int_{x_-}^{x_+} \sqrt{V(x) - V_{\min}} dx$$

is the Agmon distance between the two wells.

We would emphasize that the eigenvectors  $\psi_{\pm}$  are asymptotically localized on both wells because of the symmetry property (12). The effect on the ground state of a small perturbation  $W(x)$  that breaks the symmetry is worth mentioning. In such a case the property (12) does not work and, even if the small perturbation  $W(x)$  is supported only on one side and far from the bottom of the well, the ground state, instead of being asymptotically supported on both wells, may be localized on just one well. According to Barry Simon we may state that *the perturbation  $W(x)$  is a small flea on the elephant  $V(x)$ . The flea does not change the shape of the elephant* – in the sense that the splitting is still exponentially small – *but it can irritate the elephant enough so that it shifts its weight* – in the sense that the ground state is localized on just one well.

### Semiclassical Approximation in Any Dimension

Here we consider the eigenvalue problem (2) in any dimension  $n \geq 1$  where the potential  $V$  is assumed to be a *multi-well potential* (Dimassi and Sjöstrand 1999; Helffer 1988; Helffer and Sjöstrand 1984; Simon 1983). More precisely, we assume that

$$V_{\min} := \inf V(x) < V_{\infty} = \lim_{|x| \rightarrow +\infty} \inf V(x)$$

and for any  $E \in (V_{\min}, V_{\infty})$  we can write the decomposition of

$$V^{-1}((-\infty, E]) = \cup_{j=1}^N U_j$$

as the union of  $N$  disjoint, compact and connected sets  $U_j$ . Inside these sets, named *wells*, the classical motion is allowed; outside the classical motion is forbidden.

From well-known results we have that for energies in the interval  $(V_{\min}, V_{\infty})$  the eigenvalue problem (2) admits only discrete spectrum, that is we have only isolated eigenvalues with finite multiplicity. In order to compute these eigenvalues we'll consider, at first, the solutions of Eq. (2)

inside any single well and then we'll take into account the tunneling effect among the adjacent wells as a perturbation.

Actually, it is also possible to consider the case where  $E > V_{\infty}$ . In such a case we don't expect to have isolated eigenvalues but, under some circumstances, resonant states (Grecchi et al. 1996; Helffer and Sjöstrand 1986; Hislop and Sigal 1996). However, we don't fix our attention here on this problem.

### Semiclassical Eigenvalues at the Bottom of a Well

Let  $x_0$  be a local nondegenerate minimum for the potential  $V(x)$ . For the sake of definiteness we assume that  $x_0 = 0$  and the local minimum is such that  $V(0) = \nabla V(0) = 0$  and the Hessian matrix  $\text{Hess} = (\partial^2 V(0)/\partial x_i \partial x_j)_{i,j=1,\dots,n}$  is positively defined with eigenvalues  $2\lambda_j > 0, j = 1, \dots, n$ ; furthermore, we choose the system of coordinates such that the Hessian matrix has diagonal form, i.e.  $\text{Hess} = \text{diag}(2\lambda_j)$ .

In order to compute the eigenvalues at the bottom of the well containing the minimum  $x_0$  we approximate the potential by means of the  $n$ -dimensional harmonic oscillator. Thus, Eq. (2) takes the form

$$\sum_{j=1}^n \left[ -\hbar^2 \frac{\partial^2 \psi}{\partial x_j^2} + \lambda_j x_j^2 - E \right] \psi = 0$$

which has exact eigenvalues

$$\hbar \left[ \sum_{j=1}^n \sqrt{\lambda_j} (2\alpha_j + 1) \right], \quad \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n,$$

with normalized eigenvectors

$$(\hbar\pi)^{-n/4} \prod_{j=1}^n (2^{\alpha_j} \alpha_j!)^{-1/2} \lambda_j^{1/8} \cdot e^{-\sqrt{\lambda_j} x_j^2 / 2\hbar} H_{\alpha_j} \left( \lambda_j^{1/4} \hbar^{-1/2} x \right),$$

where  $H_m$  is the Hermite polynomial of degree  $m$ .

It is clear that when  $\hbar$  is small enough then the first energy level of the harmonic oscillator is very



close to the bottom of the potential and thus we expect that such an approximation gives the leading term of the first energy level and wavefunction of Eq. (2). In fact, there exist two formal power series

$$\begin{aligned} E(\hbar) &\sim \sum_{j=0}^{\infty} \hbar^j e_j \quad \text{and} \quad a(x; \hbar) \\ &\sim \sum_{r=0}^{\infty} \hbar^r a_r(x), \end{aligned} \quad (13)$$

where  $e_0 = V_{\min} = 0$ ,  $\hbar e_1 = \hbar \sum_{j=1}^n \sqrt{\lambda_j}$  is the first eigenvalue of the associated harmonic oscillator and  $a_0 = \left[ \prod_{j=1}^n \lambda_j / \pi^2 \right]^{1/8}$ , such that the function

$$\psi(x; \hbar) = \hbar^{-n/4} a(x; \hbar) e^{-\varphi(x)/\hbar} \quad (14)$$

is such that

$$-\hbar^2 \Delta \psi + [V(x) - E(\hbar)] \psi = O(\hbar^\infty) e^{-\varphi(x)/\hbar}$$

in a neighborhood of  $x_0 = 0$ , where  $\varphi(x) = \frac{1}{2} \sum_{j=1}^n \sqrt{\lambda_j} x_j^2 + O(|x|^2)$  is the positive solution of the Eikonal equation

$$|\nabla \varphi|^2 = V - e_0$$

in a neighborhood of  $x_0 = 0$ . The way to obtain this result essentially consists of inserting the formal power series (14) in Eq. (2) and expanding in powers of  $\hbar$  the coefficients of  $e^{-\varphi(x)/\hbar}$ , then of requiring the cancelation of the coefficients of  $\hbar^j$  for any  $j \in \mathbb{N}$ .

These approximate solutions are valid in a neighborhood of the nondegenerate minimum and they are the natural candidates to become the true eigenvalue and eigenfunction of the problem (2). In fact, we'll extend this single well approximate solution (14) to a larger domain and then we'll solve the *connection* problem.

To this end we fix an open, sufficiently small, regular bounded set  $\Omega$  containing the minimum point  $x_0$ ; then Eq. (2) with Dirichlet boundary condition on  $\Omega$ , that is

$$\psi_\Omega|_{\partial\Omega} = 0, \quad \|\psi_\Omega\|_{L^2(\Omega)} = 1, \quad (15)$$

admits one simple eigenvalue  $E_\Omega(\hbar)$  at the bottom of the well, close to the first eigenvalue of the harmonic oscillator  $e_0 + \hbar e_1$ , which admits the asymptotic expansion (13).

Actually, such an eigenvalue will depend on the choice of the domain  $\Omega$ . More precisely, if  $E_{\Omega'}(\hbar)$  denotes the first eigenvalue of Eq. (2) with Dirichlet boundary condition on  $\Omega'$  for a different domain  $\Omega'$  containing  $x_0$ , then  $E_{\Omega'}(\hbar)$  differs from  $E_\Omega(\hbar)$  for an exponentially small term. Furthermore, it is also possible to prove that, modulo an exponentially small error, the spectrum of the Schrödinger equation (2) in some interval depending on  $\hbar$  is the same as the spectrum of the direct sum of the Dirichlet single well problems in the same interval.

### Agmon Metric

In order to study the exponential behavior of the solution (14) when  $x$  is far from the minimum  $x_0$  we introduce now the Agmon distance between two points  $x_1, x_2 \in \mathbb{R}^n$  defined as

$$d_E(x_1, x_2) = \inf_{\gamma} \int_{\gamma} \sqrt{[V(x) - E]_+} dx,$$

where

$$[V(x) - E]_+ = \max \{V(x) - E, 0\},$$

$\gamma$  is any regular path connecting the two points  $x_1$  and  $x_2$  and  $E \geq V_{\min}$  is fixed. We underline that the Agmon distance defines a (pseudo)-metric on the Euclidean space  $\mathbb{R}^n$ . Indeed, it satisfies the triangle inequality

$$\begin{aligned} d_E(x_1, x_2) &\leq d_E(x_1, x_3) \\ &\quad + d_E(x_3, x_2), \quad \forall x_1, x_2, x_3 \in \mathbb{R}^n, \end{aligned}$$

but it is degenerate on each well  $U$  where  $V < E$ , in particular  $d_E(x_1, x_2) = 0$  for any couple of points  $x_1$  and  $x_2$  belonging to the same well and the *diameter* of each well  $U$  is zero with respect to this metric.

This kind of (pseudo-)metric has been used to obtain precise exponential decay of wavefunctions of Schrödinger operators. Indeed, the exponential term  $\varphi(x)$  in Eq. (14) is simply given by

$$\varphi(x) = d_{e_0}(x_0, x)$$

for any  $x$  in a neighborhood of  $x_0$ .

Furthermore, by means of the Agmon metric we are also able to give an estimate of the exponential decay of wavefunctions not only for  $x$  in a neighborhood of the minimum  $x_0$ , but also for  $x$  far from this point. In particular, let  $V_{\min} < E < V_{\infty}$  be fixed and let  $U$  be one of the wells, we consider now the eigenvalue problem for the Schrödinger equation with Dirichlet boundary conditions on a regular open set  $\Omega$  containing the well  $U$ . Let  $E_{\Omega} < E$  be an eigenvalue of Eq. (2) with Dirichlet boundary condition (15) with associated eigenfunction  $\psi_{\Omega}$ . Then, the Agmon theorem enables us to obtain the following decay estimate for the eigenfunction  $\psi_{\Omega}$ : for any fixed and positive  $\epsilon$  then

$$\left\| \nabla \left[ e^{d_E(x, U)/\hbar} \psi_{\Omega} \right] \right\|_{L^2(\Omega)} + \left\| e^{d_E(x, U)/\hbar} \psi_{\Omega} \right\|_{L^2(\Omega)} \leq C_{\epsilon} e^{\epsilon/\hbar}$$

for some positive constant  $C_{\epsilon}$  depending on  $\epsilon$  (of course we expect that  $C_{\epsilon}$  will grow as  $\epsilon$  goes to zero). That is we have a good a priori estimate of the wavefunction  $\psi_{\Omega}$  in the weighted  $H^1(\Omega)$  space with weight  $e^{d_E(x, U)/\hbar}$ . Since  $d(x, U) = 0$  for any  $x$  belonging to the well  $U$  then it follows that the solution  $\psi_{\Omega}$  is exponentially decreasing outside the classical allowed region, as already seen in the one-dimensional problems.

### Tunneling Between Wells

In order to consider the tunneling effect among the wells  $\{U_j\}_{j=1}^N$  for any fixed  $E \in (V_{\min}, V_{\infty})$  we define

$$d_E(U_i, U_j) = \inf_{x \in U_i, y \in U_j} d_E(x, y)$$

as the Agmon distance between the two wells  $U_i$  and  $U_j$ , by construction it turns out that this

distance is strictly positive for any  $i \neq j$ . A special role will be played by the minimal distance among these wells

$$S_0 = \min_{i \neq j} d_E(U_i, U_j).$$

In fact, the discrete spectrum of the Dirichlet realization of the Schrödinger equation on the boundary of the wells give, up to error of the order  $e^{-S_0/\hbar}$ , the discrete spectrum of the original Schrödinger equation (2). In order to be more definite, let  $M_j^{S, \eta}$  be the open set containing the well  $U_j$  defined as

$$M_j^{S, \eta} = \left\{ x \in \mathbb{R}^n : d_E(x, U_j) < S \text{ and } d_E(x, U_k) > \eta, k \neq j \right\}$$

that is  $M_j^{S, \eta}$  is, essentially, a ball (with respect to the Agmon pseudo-metric) large enough centered in the well  $U_j$  where we have eliminated the points from the other wells. If necessary we can regularize the boundary of  $M_j^{S, \eta}$ . In the following we take  $\eta > 0$  small enough and  $S$  large enough, in particular we require that  $S > 2S_0$ . Let  $\sigma_j$  be the discrete spectrum of the Schrödinger equation (2) with Dirichlet condition on  $M_j^{S, \eta}$  and let  $s$  be the discrete spectrum of the Schrödinger equation (2). Then, for any interval  $I(\hbar) = [\alpha(\hbar), \beta(\hbar)]$ , where  $\alpha(\hbar), \beta(\hbar) \rightarrow e_0$  as  $\hbar \rightarrow 0$ , there exists a bijection

$$b : \sigma \cap I(\hbar) \rightarrow \left[ \bigcup_{j=1}^N \sigma_j \right] \cap I(\hbar)$$

such that for any  $\rho < S_0 - 2\eta$

$$|b(\lambda) - \lambda| \leq C_{\rho} e^{-\rho/\hbar} \quad (16)$$

for some  $C_{\rho} > 0$ .

Actually, this result could be improved by assuming some regularity properties on the potential  $V$  and estimate (16) can be replaced by the precise asymptotic behavior.

If we consider now the *symmetric double well* problem where  $N = 2$  and with symmetric potential, e.g.  $V(x_1, x_2, \dots, x_n) = V(-x_1, x_2, \dots, x_n)$ , then, by symmetry, we have that the two Dirichlet realizations coincide, up to the inversion  $x_1 \rightarrow -$

$x_1$ . Then  $\sigma_1 = \sigma_2$ . If we denote by  $E_1$  and  $E_2$  the first two eigenvalues of  $\sigma$  then  $E_1 < E_2$  (in fact, the first eigenvalue is always nondegenerate) and the splitting between them can be estimated as

$$\begin{aligned} |E_2 - E_1| &\leq |E_2 - b(E_2)| + |b(E_2) - b(E_1)| \\ &\quad + |b(E_1) - E_1| \\ &\leq C_\rho e^{-\rho/\hbar}, \end{aligned}$$

since  $b(E_1) = b(E_2)$ , for any  $\rho < S_0$  since  $\eta > 0$  is arbitrary.

## Propagation of Quantum Observables

So far we have restricted our analysis to the semiclassical computation of the stable states of the time-independent Schrödinger equation (2). However, there exist some relevant results for what concerns the time evolution operator  $e^{-itH/\hbar}$ , which is the formal solution of the time-dependent Schrödinger equation (1) with Hamiltonian  $H = -\hbar^2 \Delta + V$ . In other words, this result is connected to the time-evolution of quantum observables (Egorov 1971; Robert 1988). The main tool is the  $\hbar$ -pseudodifferential calculus we briefly review below.

### Brief Review of $\hbar$ -Pseudodifferential Calculus

Here, we briefly review some basic results of the semiclassical pseudodifferential (also called  $\hbar$ -pseudodifferential) calculus. We refer to the books by Folland (1988), Grigis and Sjöstrand (1994), Martinez (2002), and Robert (1987) for a detailed treatment.

In this brief review, in order to avoid some technicalities, we restrict ourselves to the simpler case of bounded potentials; however, some of the following results hold in the general case of unbounded potentials too (see, e.g., Robert 1987). To this end we consider symbols  $a(x, y, p) \in S_{3n}(\langle p \rangle^m)$  for some positive  $m$ , that is the function  $a$  defined on  $\mathbb{R}^{3n}$  depends smoothly on  $p$  and for any  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$  one has

$$|\nabla_p^\alpha a(x, y, p)| = O(\langle p \rangle^m) \quad (17)$$

uniformly. That is

$$|\nabla_p^\alpha a(x, y, p)| \leq C \langle p \rangle^m = C[1 + |p|]^{m/2}$$

for some positive constant  $C$  independent of  $x$ ,  $y$  and  $p$ , where  $\nabla_p^\alpha = \left( \partial_{p_1}^{\alpha_1}, \dots, \partial_{p_n}^{\alpha_n} \right)$  and  $|p| = |p_1| + \dots + |p_n|$ .

We associate to a symbol  $a \in S_{3n}(\langle p \rangle^m)$  the semiclassical pseudodifferential operator of degree  $m$  defined for  $u \in C_0^\infty(\mathbb{R}^n)$  as the Fourier integral operator

$$\begin{aligned} [\text{Op}_h(a)u](x) &= \frac{1}{[2\pi\hbar]^n} \\ &\cdot \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i(x-y) \cdot p/\hbar} a(x, y, p) u(y) dy dp \end{aligned}$$

We notice that it is formally self-adjoint on the Hilbert space  $L^2(\mathbb{R}^n)$  when the symbol  $a$  is such that  $a(x, y, p) = \overline{a(y, x, p)}$ .

The above pseudodifferential operator can be extended in a unique way to a linear continuous operator on the space of smooth rapidly decreasing functions  $\mathcal{S}'(\mathbb{R}^n)$  and its dual space of tempered distributions  $\mathcal{S}'(\mathbb{R}^n)$ . Furthermore, when the symbol  $a \in S_{3n}(1)$  then the associated pseudodifferential operator is continuous on  $L^2(\mathbb{R}^n)$  (this result is the so-called Calderón–Villancourt theorem).

It is easy to see that this class of pseudodifferential operators contains the usual differential operators. For instance, in the particular case where the symbol  $a$  has the form

$$a(x, y, p) = \sum_{|\alpha| \leq m} b_\alpha(x) p^\alpha$$

then its associated operator is the differential operator given by

$$\begin{aligned} [\text{Op}_h(a)u](x) &= \left[ \text{Op}_h \left( \sum_{|\alpha| \leq m} b_\alpha(x) p^\alpha \right) u \right](x) \\ &= \sum_{|\alpha| \leq m} b_\alpha(x) (i\hbar \nabla_x)^\alpha u(x). \end{aligned}$$

The class of pseudodifferential operator is closed with respect to the composition. That is: given two symbols  $a \in S_{3n}(\langle p \rangle^m)$  and

$b \in S_{3n}(\langle p \rangle^{m'})$  then the composition of the associated pseudodifferential operators is still a pseudodifferential operator:

$$\text{Op}_h(a) \circ \text{Op}_h(b) = \text{Op}_h(c)$$

where the symbol  $c \in S_{3n}(\langle p \rangle^{m+m'})$  depends on  $\hbar$  and it is given by the *Moyal product*

$$\begin{aligned} c(x, y, p) &= (a \# b)(x, y, p) = \frac{1}{[2\pi\hbar]^n} \\ &\cdot \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i(x-z) \cdot (\eta-p)/\hbar} a(x, z, \eta) b(z, y, p) dz d\eta \\ &\sim \sum_{|\alpha| \geq 0} \frac{\hbar^{|\alpha|}}{i^{|\alpha|} \alpha!} \nabla_z^\alpha \nabla_\eta^\alpha [a(x, z, \eta) b(z, y, p)] \Big|_{z=x, \eta=p} \end{aligned}$$

as  $\hbar$  goes to zero; we should underline that the above asymptotic formula becomes an exact formula when the symbol  $a$  is polynomial with respect to  $p$  since the sum becomes finite. As a result it follows that for any *elliptic symbol*  $a \in S_{3n}(\langle p \rangle^m)$ , that is such that  $|a(x, y, p)| \geq C \langle p \rangle^m$  for some positive constant  $C$ , then its associate pseudodifferential operator is invertible in the sense that there exists a symbol  $b \in S_{3n}(\langle p \rangle^{-m})$ , depending on  $\hbar$ , where

$$\text{Op}_h(a) \circ \text{Op}_h(b) = 1 + \text{Op}_h(r)$$

and

$$\text{Op}_h(b) \circ \text{Op}_h(a) = 1 + \text{Op}_h(s),$$

with  $r, s = O(\hbar^\infty)$  in  $S_{3n}(1)$ . The symbol  $b \sim \sum_j \hbar^j b_j$  is iteratively defined where  $b_0 = \frac{1}{a}$ .

Now, we are ready to define the *quantization of a classical observable*. Classical observables are functions  $a(x, p)$  of the position  $x \in \mathbb{R}^n$  and of the momenta  $p \in \mathbb{R}^n$ , if we assume that  $a$  is a smooth function such that

$$|\nabla_p^\alpha a(x, p)| \leq C \langle p \rangle^m$$

then, for any  $\rho \in [0, 1]$ , it follows that

$$a^\rho(x, y, z) = a[(1 - \rho)x + \rho y, p] \in S_{3n}(\langle p \rangle^m).$$

We define the *quantization of the observable a* as:

$$\text{Op}_h^\rho(a) = \text{Op}_h(a^\rho)$$

where the values  $\rho = 0, \frac{1}{2}, 1$  will play a special role; in particular, for  $\rho = 0$  we have the *standard* (also called *left*) quantization, for  $\rho = 1$  we have the *right* quantization and for  $\rho = \frac{1}{2}$  we have the *Weyl* quantization

$$\text{Op}_h^W(a) := \text{Op}_h^{1/2}(a) = \text{Op}_h(a^{1/2}).$$

We emphasize that the Weyl quantization is particularly important in quantum mechanics because when the classical observable  $a$  is a real valued function then the associate pseudodifferential operator  $\text{Op}_h^W(a)$  is formally self-adjoint on  $L^2(\mathbb{R}^n)$ .

We close this review by recalling the following important result which connects the commutator  $[A, B] = AB - BA$  of the pseudodifferential operators  $A = \text{Op}_h^\rho(a)$  and  $B = \text{Op}_h^\rho(b)$ , where  $a, b$  are two classical observables, and the *Poisson bracket* of  $a$  and  $b$ : let  $c$  be the unique symbol such that  $\text{Op}_h^\rho(c) = [A, B]$ , then

$$c = -i\hbar\{a, b\} + O(\hbar^2).$$

### Egorov Theorem

Now, we are ready to compare the classical evolution of a classical observable with the quantum evolution of its quantum counterpart. With more details let  $h(x, p)$  be a classical Hamiltonian where  $x \in \mathbb{R}^n$  denotes the spatial variable and  $p \in \mathbb{R}^n$  denotes the momentum. Let

$$\phi^t : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$$

be the (classical) Hamiltonian flux. Thus, for any classical observable function  $b = b(x, p) \in C^\infty(\mathbb{R}^{2n}; \mathbb{R})$  let

$$b^t(x, p) = (b \circ \phi^t)(x, p) = b[\phi^t(x, p)]$$

be the classical evolution of this observable.

It is well known that we can associate to any real-valued classical observable  $b(x, p)$  a symmetric  $\hbar$ -pseudodifferential linear operator denoted by  $\text{Op}_\hbar^W(b)$  by means of the semiclassical (Weyl) quantization rule formally defined as

$$[\text{Op}_\hbar^W(b)u](x) := \frac{1}{[2\pi\hbar]^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i(x-y)p/\hbar} b\left(\frac{x+y}{2}, p\right) u(y) dy dp$$

where  $(x-y) \cdot p = \sum_{i=1}^n (x_i - y_i) p_i$ . In order to properly define this integral operator on the Hilbert space  $L^2(\mathbb{R}^n)$  we require that estimate (17) holds; actually, to the present purposes it is sufficient to assume the following weaker condition on the observable  $b$ :

$$|\nabla_{x,p}^\alpha b(x, p)| \leq C [1 + |x|^2 + |p|^2]^{m/2} \quad (18)$$

for some  $m \geq 0$  and any  $\alpha \in \mathbb{N}^{2n}$ .

For instance, if  $h_0$  is the Hamiltonian associated to a harmonic oscillator, then it is a quadratic function with respect to both position and momentum variables

$$h_0(x, p) = \sum_{j=1}^n [p_j^2 + \omega_j^2 x_j^2] \quad (19)$$

and the associated Hamiltonian operator (let  $n = 1$  for the sake of simplicity) takes the usual form

$$\begin{aligned} [H_0 u](x) &= [\text{Op}_\hbar^W(h_0)u](x) \\ &= \frac{1}{2\pi\hbar} \int_{\mathbb{R} \times \mathbb{R}} e^{i(x-y)p/\hbar} \left[ p^2 + \omega^2 \left( \frac{x+y}{2} \right)^2 \right] \\ &\quad \cdot u(y) dy dp \\ &= -\hbar^2 \frac{d^2}{dx^2} + \omega^2 x^2 \end{aligned}$$

by integrating by parts twice and since  $1/(2\pi\hbar) \int_{\mathbb{R}} e^{i(x-y)p/\hbar} dp = \delta(x-y)$ .

In such a way it is possible to associate a classical observable  $b$  to a quantum operator  $B$ . If we denote by  $B = \text{Op}_\hbar^W(b)$  and  $H = \text{Op}_\hbar^W(h)$

the operators associated to the classical observable  $b$  and to the Hamiltonian  $h$  then the time quantum evolution  $B^t = e^{itH/\hbar} B e^{-itH/\hbar}$  of the observable  $B$  solves the Heisenberg equation

$$\frac{dB^t}{dt} = \frac{i}{\hbar} [H, B^t]$$

and it is, in some sense, related to the classical evolution  $b^t$  as we are going to explain.

In the very particular case where the Hamiltonian  $h_0$  is given by the harmonic oscillator (19) then we have that the quantum evolution  $B^t$  of the observable  $B$  is a  $\hbar$ -pseudodifferential operator with symbol  $b^t$  given by means of the classical flux Hamiltonian; that is

$$e^{itH_0/\hbar} \text{Op}_\hbar^W(b) e^{-itH_0/\hbar} = \text{Op}_\hbar^W(b \circ \phi^t)$$

where  $\phi^t$  is the Hamiltonian flux generated by the Hamiltonian (19).

In other words, *for the harmonic oscillator Hamiltonian then quantum evolution and Weyl quantization commute*. This property is specific for this problem and it comes from the fact that the flux Hamiltonian is a linear function with respect to  $(x, p)$ . This fact is not true in a general case. However, it is possible to see that a generalization of such a result holds when  $H_0$  is replaced by any symmetric  $\hbar$ -pseudodifferential operator  $H$ ; this result is the so-called Egorov theorem (see Egorov (1971) for the original statement, see also Robert (1988) for a detailed review). In particular, under some assumptions, it is possible to prove that the zero-th order remainder term

$$R(t) := e^{itH/\hbar} \text{Op}_\hbar^W(b) e^{-itH/\hbar} - \text{Op}_\hbar^W(b \circ \phi^t)$$

is a bounded operator such that  $\|R(t)\| = O(\hbar)$  in the sense that the norm of the operator  $R(t)$  is bounded

$$\|R(t)\| \leq C\hbar$$

for some  $C > 0$  and for any  $t \geq 0$ .

Furthermore, it is possible to extend this asymptotic result to any order  $N \geq 1$  for a suitable choice of  $p_n$ , that is the classical and quantum

evolution coincides up to a term of order  $\hbar^{N+1}$  in the semiclassical limit:

$$\begin{aligned} e^{itH/\hbar} \text{Op}_h^W(b) e^{-itH/\hbar} - \text{Op}_h^W(b \circ \phi^t) \\ - \sum_{n=1}^N \hbar^n \text{Op}_h^W(p_{n,t}) \\ = O(\hbar^{N+1}) \end{aligned}$$

in the norm sense.

## Future Directions

Semiclassical methods are a field of research where theoretical results are rapidly evolving. Just to name some active research topics:  $\hbar$ -pseudodifferential operators, Weyl functional calculus, frequency sets, semiclassical localization of eigenfunctions, semiclassical resonant states, Born–Oppenheimer approximation, stability of matter and Scott conjecture, semiclassical Lieb–Thirring inequality, Peierls substitution rule, etc. Furthermore, semiclassical methods have been also successfully applied in different contexts such as superfluidity and statistical mechanics.

Looking forward, we see new emerging research fields in the area of semiclassical methods: *numerical WKB interpolation techniques* and *semiclassical nonlinear Schrödinger equations*.

Indeed, the recent researches in the area of nanosciences and nanotechnologies have opened up new fields where models for semiconductor devices of increasingly small size and electric charge transport along nanotubes cannot be fully understood without considering their quantum nature. Since the oscillating behavior of the solutions of the Schrödinger equation induces serious difficulties for standard numerical simulations, then new numerical approaches based on WKB interpolation are required (Ben Abdallah and Pinaud 2006; Bonnaillie-Noël et al. 2006; Presilla and Sjöstrand 1996).

Although the nonlinear Schrödinger equation has been an argument of theoretical research since the 1970s, only in the last few years, with the successful experiments on Bose–Einstein

condensate states, has an increasing interest been shown. When we add a nonlinear term to the time-dependent Schrödinger equation (1) then the dynamics of the model drastically changes and new peculiar features, such as the *blow-up effect* and *stability of stationary states*, appear. Semiclassical arguments applied to nonlinear Schrödinger equations justify their reduction to finite dimensional dynamical systems and thus it is possible to obtain an approximate solution, at least for nonlinear time-dependent Schrödinger equations in small dimensional spaces, typically for  $n = 1$  and  $n = 2$ . The extension of this technique to the case  $n > 2$  and the validity of such an approximation for large times is still an open problem (Grecchi et al. 2002; Raghavan et al. 1999; Sacchetti 2005).

## Bibliography

- Ben Abdallah N, Pinaud O (2006) Multiscale simulation of transport in an open quantum system: resonances and WKB interpolation. *J Comp Phys* 213:288–310
- Berezin FA, Shubin MA (1991) *The Schrödinger equation*. Kluwer, Dordrecht
- Berry MV, Mount KE (1972) Semiclassical approximation in wave mechanics. *Rep Prog Phys* 35:315–397
- Bonnaillie-Noël V, Nier F, Patel Y (2006) Computing the steady states for an asymptotic model of quantum transport in resonant heterostructures. *J Comp Phys* 219: 644–670
- Claviere P, Jona Lasinio G (1986) Instability of tunneling and the concept of molecular structure in quantum mechanics: the case of pyramidal molecules and the enantiomer problem. *Phys Rev A* 33:2245–2253
- Dimassi M, Sjöstrand J (1999) *Spectral asymptotics in the semi-classical limit*, London Mathematical Society lecture note series 268. Cambridge University Press, Cambridge, UK
- Dingle RB (1973) *Asymptotic expansion: their derivation and interpretation*. Academic, London
- Egorov YV (1971) Canonical transformation of pseudodifferential operators. *Trans Moscow Math Soc* 24:1–24
- Folland G (1988) *Harmonic analysis in phase space*. Princeton University Press, Princeton
- Fröman N, Fröman PO (1965) JWKB approximation. North-Holland, Amsterdam
- Fröman N, Fröman PO (2002) *Physical problems solved by the phase-integral methods*. Cambridge University Press, Cambridge, UK
- Graffi S, Grecchi V, Jona-Lasinio G (1984) Tunneling instability via perturbation theory. *J Phys A Math Gen* 17:2935–2944

- Grecchi V, Martinez A, Sacchetti A (1996) Splitting instability: the unstable double wells. *J Phys A Math Gen* 29:4561–4587
- Grecchi V, Martinez A, Sacchetti A (2002) Destruction of the beating effect for a non-linear Schrödinger equation. *Commun Math Phys* 227:191–209
- Grigis B, Sjöstrand J (1994) Microlocal analysis for differential operators. An introduction, London Mathematical Society lecture note series 196. Cambridge University Press, Cambridge, UK
- Harrell EM (1980) Double wells. *Commun Math Phys* 75: 239–261
- Helffer B (1988) Semi-classical analysis for the Schrödinger operator and applications, Lecture notes in mathematics 1336. Springer, Berlin
- Helffer B, Sjöstrand J (1984) Multiple wells in the semiclassical limit I. *Commun Partial Differ Equ* 9: 337–408
- Helffer B, Sjöstrand J (1986) Resonances en limite semi-classique. *Mém Soc Math France* 24–25:1–228
- Hislop P, Sigal IM (1996) Introduction to spectral theory, Applied mathematical sciences, vol 113. Springer, New York
- Landau LD, Lifshitz EM (1959) Quantum mechanics, Course of theoretical physics. Pergamon, Oxford, UK
- Martinez A (2002) An introduction to semiclassical and microlocal analysis. Springer, New York
- McHugh JAM (1971) An historical survey of ordinary linear differential equations with a large parameter and turning points. *Arch Hist Exact Sci* 7: 277–324
- Merzbacher E (1970) Quantum mechanics, 2nd edn. Wiley, New York
- Olver FWJ (1974) Asymptotics and special functions. Academic, New York
- Presilla C, Sjöstrand J (1996) Transport properties in resonant tunneling heterostructures. *J Math Phys* 37:4816–4844
- Raghavan S, Smerzi A, Fantoni S, Shenoy SR (1999) Coherent oscillations between two weakly coupled Bose–Einstein condensates: Josephson effects, p oscillations, and macroscopic quantum self-trapping. *Phys Rev A* 59:620–633
- Robert D (1987) Autour de l'Approximation Semi-classique. Birkhäuser, Basel
- Robert D (1988) Semi-classical approximation in quantum mechanics. A survey of old and recent mathematical results. *Helv Phys Acta* 71:44–116
- Sacchetti A (2005) Nonlinear double well Schrödinger equations in the semiclassical limit. *J Stat Phys* 119: 1347–1382
- Simon B (1983) Semiclassical limit of low lying eigenvalues I: non degenerate minima. *Ann IHP Phys Theor* 38:295–307
- Simon B (1985) Semiclassical limit of low lying eigenvalues IV: the flea of the elephant. *J Funct Anal* 63: 123–136
- Voros A (1982) Spectre de l'Équation de Schrödinger et Méthode BKW. Publications Mathématiques d'Orsay 81.09
- Wilkinson M, Hannay JH (1987) Multidimensional tunneling between excited states. *Phys D Nonlinear Phenom* 27:201–212





# Perturbation Theory and Molecular Dynamics

Gianluca Panati

Dipartimento di Matematica, Università di Roma  
“La Sapienza”, Rome, Italy

## Article Outline

Glossary

Definition of the Subject

Introduction

The Framework

The Leading Order Born–Oppenheimer  
Approximation

Beyond the Leading Order

Future Directions

Bibliography

## Glossary

**Adiabatic decoupling** In a complex system (either classical or quantum), the dynamical decoupling between the slow and the fast degrees of freedom.

**Adiabatic perturbation theory** A mathematical algorithm which exploits the adiabatic decoupling of degrees of freedom in order to provide an approximated (but yet accurate) description of the slow part of the dynamics. In the framework of QMD, it is used to approximately describe the dynamics of nuclei, the perturbative parameter  $\varepsilon$  being related to the small electron/nucleus mass ratio.

**Electronic structure problem** The problem consisting in computing, at fixed positions of the nuclei, the energies (eigenvalues) and eigenstates corresponding to the electrons. An approximate solution is usually obtained numerically.

**Molecular dynamics** The dynamics of the nuclei in a molecule. While a first insight in the problem can be obtained by using classical mechanics (Classical Molecular Dynamics), a

complete picture requires quantum mechanics (Quantum Molecular Dynamics) Perturbation Theory in Quantum Mechanics. This contribution focuses on the latter viewpoint.

## Definition of the Subject

In the framework of Quantum Mechanics the dynamics of a molecule is governed by the (time-dependent) Schrödinger equation, involving nuclei and electrons coupled through electromagnetic interactions. While the equation is mathematically well-posed, yielding the existence of a unique solution, the complexity of the problem makes the exact solution unattainable. Even for small molecules, the large number of degrees of freedom prevents from *direct* numerical simulation, making an approximation scheme necessary.

Indeed, one may exploit the smallness of the electron/nucleus mass ratio to introduce a convenient computational scheme leading to approximate solutions of the original time-dependent problem. In this article we review the standard approximation scheme (**dynamical Born–Oppenheimer approximation**) together with its ramifications and some recent generalizations, focusing on mathematically rigorous results.

The success of this approximation scheme is rooted in a clear separation of time-scales between the motion of electrons and nuclei. Such separation provides the prototypical example of **adiabatic decoupling** between the fast and the slow part of a quantum dynamics. More generally, adiabatic separation of time-scales plays a fundamental role in the understanding of complex systems, with applications to a wide range of physical problems.

## Introduction

Through the discovery of the Schrödinger equation the theoretical physics and chemistry community attained a powerful tool for computing atomic spectra, either exactly or in perturbation

expansion. Born and Oppenheimer (1927) immediately strived for a more ambitious goal, namely to understand the excitation spectrum of molecules on the basis of the new wave mechanics. They accomplished to exploit the small electron/nucleus mass ratio as an expansion parameter, which then leads to the **stationary** Born–Oppenheimer approximation. Since then it has become a standard and widely used tool in quantum chemistry, now supported by rigorous mathematical results (Combes 1977; Combes et al. 1981; Hagedorn 1987, 1988a; Klein et al. 1992).

Beyond molecular structure and excitation spectra, dynamical processes have gained in interest. Examples are the scattering of molecules, chemical reactions, or the decay of an excited state of a molecule through tunneling processes. Such problems require a **dynamical** version of the Born–Oppenheimer approximation, which is the topic of this article. At the leading order, the electronic energy at fixed positions of the nuclei serves as an effective potential between the nuclei. We call this the zeroth order Born–Oppenheimer approximation. The resulting effective Schrödinger equation can be used for both statical and dynamical purposes. The input is an electronic structure calculation, which for the purpose of our article we regard as given by other means.

While there are many physical and chemical properties of molecules explained by the zeroth order Born–Oppenheimer approximation, there are cases where higher order corrections are required. Famous examples are the dynamical Jahn–Teller effect and the dynamics of singled out nuclear degrees of freedom near the conical intersection of two energy surfaces. The first order Born–Oppenheimer approximation involves geometric phases, which are of great interest also in other domains of Quantum Mechanics ((Weigert and Littlejohn 1993), “Quantum Bifurcations”).

Finally, we mention that some dynamical processes can be modeled as scattering problems. In such cases it is convenient to combine the Born–Oppenheimer scheme together with scattering theory, a topic which goes beyond the purpose of this contribution (see (Klein et al. 1993, 1997) and references therein).

A complete overview of the vast literature on the subject of the dynamical Born–Oppenheimer approximation is provided in (Hagedorn and Joye 2007).

## The Framework

We consider a molecule consisting of  $K$  nuclei, whose positions are denoted as  $x = (x_1, \dots, x_K) \in \mathbb{R}^{3K} =: X$ , and  $N$  electrons, with positions  $y = (y_1, \dots, y_N) \in \mathbb{R}^{3N} =: Y$ . The wavefunction of the molecule is therefore a square-integrable function  $\Psi$  depending on all these coordinates.

Molecular dynamics is described through the Schrödinger equation

$$1\hbar \frac{d}{ds} \Psi_s = H_{\text{mol}} \Psi_s \quad (1)$$

where  $s$  denotes time measured in microscopic units and the Hamiltonian operator is given by

$$H_{\text{mol}} = - \sum_{k=1}^K \frac{\hbar^2}{2M_k} \Delta_{x_k} - \sum_{i=1}^N \frac{\hbar^2}{2m_e} \Delta_{y_i} + V_e(y) + V_n(x) + V_{\text{en}}(x, y). \quad (2)$$

Here  $\hbar$  is the Planck constant,  $m_e$  is the mass of the electron and  $M_k$  the mass of the  $k$ th nucleus, and the interaction terms are explicitly given by

$$V_n(x) = \sum_{k=1}^K \sum_{l \neq k}^K \frac{e^2 Z_k Z_l}{|x_k - x_l|},$$

$$V_e(y) = \sum_{i=1}^N \sum_{j \neq i}^N \frac{e^2}{|y_i - y_j|},$$

and

$$V_{\text{en}}(x, y) = \sum_{k=1}^K \sum_{i=1}^N - \frac{e^2 Z_k}{|x_k - y_i|},$$

where  $eZ_k$ , for  $Z_k \in \mathbb{Z}$ , is the electric charge of the  $k$ th nucleus. In some cases, to obtain rigorous mathematical results one needs to slightly smear out the charge distribution of the nuclei. This is in

agreement with the physical picture that nuclei are not point like but extended objects. Hereafter we will assume, for sake of a simpler notation, that all the nuclei have the same mass  $M$ . The subsequent discussion is still valid in the general case, with the appropriate choice of the adiabatic parameter indicated below.

As mentioned above, the large number of degrees of freedom makes convenient to elaborate an approximation scheme, exploiting the smallness of the parameter

$$\varepsilon := \sqrt{\frac{m_e}{M}} = 10^{-2} \dots 10^{-3}. \quad (3)$$

In the general case, one has to choose  $\varepsilon = \max \left\{ \sqrt{m_e/M_k} : 1 \leq k \leq K \right\}$ .

By introducing atomic units ( $\hbar = 1, m_e = 1$ ) and making explicit the role of the **adiabatic parameter**  $\varepsilon$ , the Hamiltonian  $H_{\text{mol}}$  reads (up to a change of energy scale)

$$H_\varepsilon = - \sum_{k=1}^K \frac{\varepsilon^2}{2} \Delta_{x_k} + V_n(x) + \underbrace{\sum_{i=1}^N -\frac{1}{2} \Delta_{y_i} + V_e(y) + V_{\text{en}}(x, y)}_{H_{\text{el}}(x)}. \quad (4)$$

Notice that, for each fixed nuclei configuration  $x = (x_1, \dots, x_K) \in X$ , the operator  $H_{\text{el}}(x)$  is an operator acting on the Hilbert space  $\mathcal{H}_{\text{el}}$  corresponding to the electrons alone.

If the kinetic energies of the nuclei and the electrons are comparable, as it happens in the vast majority of physical situations in view of energy equipartition, then the velocities scale as

$$|v_n| \approx \sqrt{\frac{m_e}{M}} |v_e| = \varepsilon |v_e|, \quad (5)$$

where  $v_n$  and  $v_e$  denotes respectively the typical velocity of nuclei and electrons. Therefore, in order to observe a non-trivial dynamics for the nuclei, one has to wait a microscopically long time, namely a time of order  $O(\varepsilon^{-1})$ . This scaling fixes the **macroscopic time scale**  $t$ , together with

the relation  $t = \varepsilon s$ , where  $s$  is the microscopic time appearing in Eq. (1). We are therefore interested in the behavior of the solutions of the equation

$$1\varepsilon \frac{d}{dt} \Psi(t) = H_\varepsilon \Psi(t) \quad (6)$$

in the limit  $\varepsilon \rightarrow 0$ .

We assume as an input a solution of the **electronic structure problem**, i.e. that for every fixed configuration of the nuclei  $x = (x_1, \dots, x_K)$  one knows the solution of the eigenvalue problem

$$H_e(x) \chi_j(x) = E_j(x) \chi_j(x), \quad (7)$$

with  $E_j(x) \in \mathbb{R}$  and  $\chi_j(x) \in \mathcal{H}_{\text{el}}$ . Since electrons are fermions, one has  $\mathcal{H}_{\text{el}} = S_a L^2(\mathbb{R}^{3N})$  with  $S_a$  projecting onto the antisymmetric wave functions. The eigenvectors in Eq. (7) are normalized as

$$\langle \chi_j(x), \chi_\ell(x) \rangle_{\mathcal{H}_{\text{el}}} \equiv \int_Y \chi_j^*(x, y) \chi_\ell(x, y) dy = \delta_{j\ell}$$

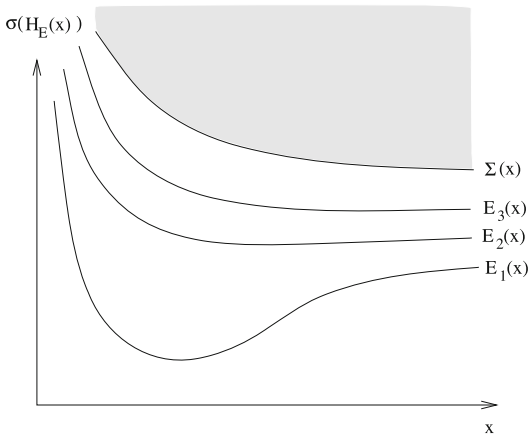
with respect to the scalar product in  $\mathcal{H}_{\text{el}}$ . Note that the eigenvectors are determined only up to a phase  $\vartheta_j(x)$ . Generically, in addition to the bound states,  $H_e(x)$  has continuous spectrum. We label the eigenvalues in Eq. (7) as

$$E_1(x) \leq E_2(x) \leq \dots, \quad (8)$$

including multiplicity. The graph of  $E_j$  is called the  **$j$ th energy surface** or energy band, see Fig. 1.

Generically, in realistic examples such energy bands cross each other, and the possible structures of band crossing have been classified (Hagedorn 1992; von Neumann and Wigner 1929). Figure 2 illustrates a realistic example of energy bands, showing in particular the typical conical intersection of two energy surfaces.

Let  $\psi(x)$  be a nucleonic wave function,  $\psi \in \mathcal{H}_n$ , the Hilbert space corresponding to the nuclei. For simplicity we take  $\mathcal{H}_n = L^2(X)$ , remembering that to impose the physically correct statistics for the nuclei requires extra considerations (Mead and Truhlar 1979). States  $\Psi$  of the molecules with the property that the electrons are precisely in the  $j$ th eigenstate are then of the form



**Perturbation Theory and Molecular Dynamics, Fig. 1** A schematic representation of energy bands. At each fixed nuclei configuration  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_K)$  the electronic Hamiltonian  $H_{\text{el}}(\mathbf{x})$  exhibits point spectrum (corresponding to states with all the electrons bound to the nuclei) and continuous spectrum (corresponding to states in which one or more electrons are quasi-free, i.e. the molecule is ionized)

$$\Psi(x, y) = \psi(x)\chi_j(x, y). \quad (9)$$

We can think of  $\Psi$  either as a wave function in the total Hilbert space  $\mathcal{H} = \mathcal{H}_n \otimes \mathcal{H}_{\text{el}}$ , or as a wave function for the electrons (i.e. an element of  $\mathcal{H}_{\text{el}}$ ) depending parametrically on  $x$ . In the common jargon, a state in the form Eq. (9) is said a state *concentrated on the  $j$ th band*. We denote as  $P_j$  the projector on the subspace consisting of states of the form Eq. (9); since the  $\{\chi_j(x)\}_j \in \mathbb{N}$  are orthonormal,  $P_j$  is indeed an orthogonal projection in  $\mathcal{H}_n \otimes \mathcal{H}_{\text{el}}$ .

## The Leading Order Born–Oppenheimer Approximation

We focus now on an a specific energy band, say  $E_n$ , assuming that it is globally isolated from the rest of the spectrum (the behavior of the wavefunction at the crossing points will be addressed later).

Under such assumption, a state  $\Psi_0$  which is initially concentrated on the  $n$ th band will stay localized in the same band up to errors of order  $O(\varepsilon)$ : more specifically, one shows that

$$\|(1 - P_n)e^{-iH_\varepsilon t/\varepsilon}P_n\Psi_0\|_{\mathcal{H}} = O(\varepsilon). \quad (10)$$

The space  $\text{Ran } P_n$ , consisting of wavefunctions in the form Eq. (9), is usually called the **adiabatic subspace** corresponding to the  $n$ th band.

Since  $\text{Ran } P_n$  is approximately invariant under the dynamics, one may wonder whether there is a simple and convenient way to approximately describe the dynamics insidesuch subspace. Indeed, one may argue that for an initial state in  $\text{Ran } P_n$  the dynamics of the nuclei is governed by the reduced Hamiltonian

$$P_n H_\varepsilon P_n = -\frac{\varepsilon^2}{2} \sum_{k=1}^K \Delta_{x_k} + V_n(x) + E_n(x) + O(\varepsilon) \quad (11)$$

acting in  $\text{Ran } P_n \cong \mathcal{H}_n = L^2(X)$ . The **dynamical Born–Oppenheimer approximation** consists, at the leading order, in replacing the original Hamiltonian Eq. (4) by the Hamiltonian

$$H_{\text{BO}} = -\frac{\varepsilon^2}{2} \sum_{k=1}^K \Delta_{x_k} + V_n(x) + E_n(x) \quad (12)$$

acting in  $L^2(X)$ , getting thus an impressive dimensional reduction. In other words: let  $\Psi(t, x, y)$  be the solution of Eq. (6) with initial datum  $\Psi_0(x, y) = \varphi_0(x)\chi_n(x, y)$ ; then  $\Psi(t, x, y) = \varphi(t, x)\chi_n(x, y) + O(\varepsilon)$  where  $\varphi(t, x)$  is the solution of the effective equation

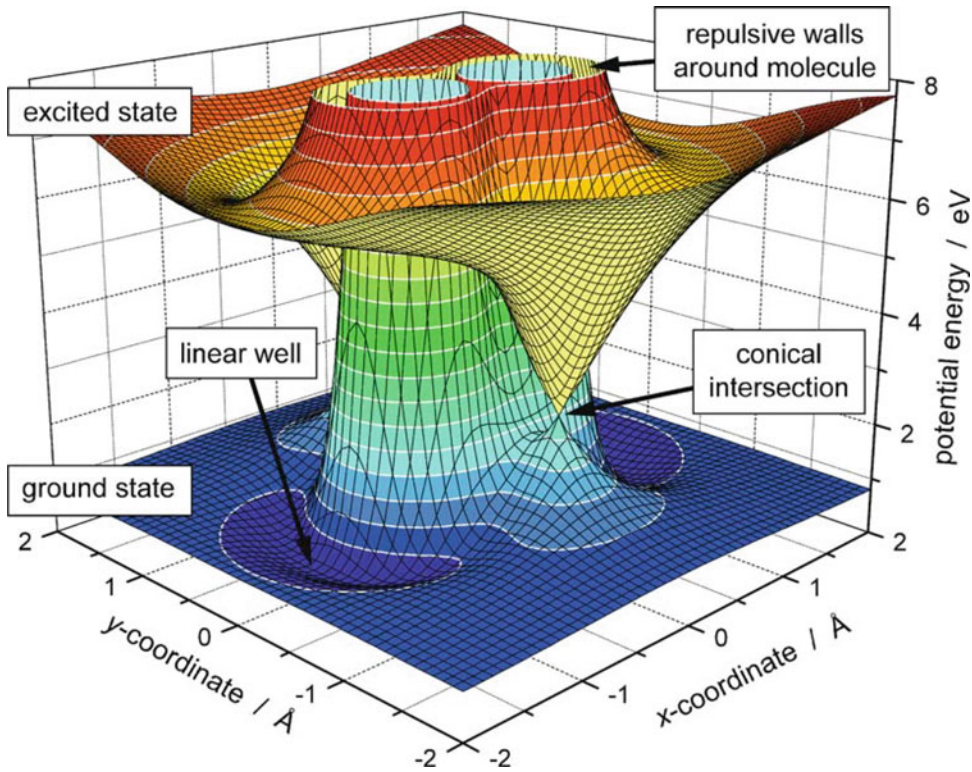
$$1 \frac{d}{dt} \varphi(t) = H_{\text{BO}} \varphi(t) \quad (13)$$

with initial datum  $\varphi_0$ .

To prove mathematically the previous claim, one has to bound the difference

$$\left( e^{-iH_\varepsilon t/\varepsilon} - e^{-iP_n H_\varepsilon P_n t/\varepsilon} \right) P_n.$$

A proof of this fact is not immediate as one might expect. Indeed, the Duhamel method (consisting essentially in rewriting a function as the integral of its derivative) yields



**Perturbation Theory and Molecular Dynamics, Fig. 2** The first two energy bands for the hydrogen quasi-molecule  $\text{H}_3$ , i.e. the system consisting of three protons and three electrons. The picture shows the restriction of such bands over a 2-dimensional subspace of the configuration space. Two hydrogen nuclei are located on the  $x$ -axis with a fixed separation of 1.044 Angstrom  $\text{\AA}$ , and

the energy bands are plotted as a function of the relative position  $(x, y)$  of the third nucleus. Notice the conical intersection between the ground and the first excited state, which appears at equilateral triangular geometries. (© Courtesy of Eckart Wrede, Durham University. The plot is generated using the analytic representation of the  $\text{H}_3$  energy bands obtained in (Varandas et al. 1987))

$$\begin{aligned}
 & \left( e^{-iH_\varepsilon t/\varepsilon} - e^{-iP_n H_\varepsilon P_n t/\varepsilon} \right) P_n \\
 &= i\varepsilon e^{-iH_\varepsilon t/\varepsilon} \int_0^{t/\varepsilon} ds \, e^{iH_\varepsilon s} (P_n H_\varepsilon P_n - H_\varepsilon) e^{-iP_n H_\varepsilon P_n s} P_n \\
 &= i\varepsilon e^{-iH_\varepsilon t/\varepsilon} \int_0^{t/\varepsilon} ds \, e^{iH_\varepsilon s} (P_n H_\varepsilon P_n - H_\varepsilon) P_n e^{-iP_n H_\varepsilon P_n s} \\
 &= i\varepsilon e^{-iH_\varepsilon t/\varepsilon} \int_0^{t/\varepsilon} ds \, e^{iH_\varepsilon s} \underbrace{[P_n, H_\varepsilon] P_n}_{O(\varepsilon)} e^{-iP_n H_\varepsilon P_n s}
 \end{aligned}$$

The commutator appearing in the last line is estimated as

$$[P_n, H_\varepsilon] P_n = \left[ |\chi_n(x)\rangle \langle \chi_n(x)|, -\frac{\varepsilon^2}{2} \Delta_x \right] P_n = O(\varepsilon), \quad (14)$$

but the integration interval diverges as  $O(\varepsilon^{-1})$ . Thus the *naïve approach* fails. A rigorous proof has been provided in (Spohn and Teufel 2001), elaborating on (Kato 1950), exploiting the fact that the integral in Eq. (14) is, roughly speaking, an oscillatory operator integral. A more direct approach, based on the evolution of generalized Gaussian wavepackets, has been followed in the pioneering papers by Hagedorn (Hagedorn 1986, 1988b).

## Beyond the Leading Order

The dynamics of a state initially concentrated on an isolated energy band is described, up to errors



of order  $O(\varepsilon)$ , by the Born–Oppenheimer dynamics Eq. (13). It is physically interesting to find an effective dynamics which approximates the original dynamics with a higher degree of accuracy. At first sight one might think that this goal can be simply reached by expanding the operator  $P_n H_\varepsilon P_n$  to the next order in  $\varepsilon$ . (Notice that the first term appearing in Eq. (11) does contribute as a term of  $O(1)$ , since we are considering states such that the kinetic energy of the nuclei is not vanishing, i.e.  $\| -\varepsilon^2 \Delta_x \Psi \| = O(1)$ , in agreement with the mentioned energy equipartition). However such *naïf* expansion has no physical meaning since it makes no sense to compute the operator appearing in Eq. (11) with greater accuracy if the space  $\text{Ran } P_n$  itself is invariant only up to terms of order  $O(\varepsilon)$ .

To get a deeper insight in the problem, one has to investigate the origin of the  $O(\varepsilon)$  term appearing in the Eq. (10): either (a) there is a part of the wavefunction of order  $O(\varepsilon)$  which is scattered in all the directions in the Hilbert space, or (b) still there is a subspace invariant up to smaller errors, which is however tilted with respect to  $\text{Ran } P_n$  by a term of order  $O(\varepsilon)$ .

Therefore, two natural questions arise:

- (i) **Almost-invariant subspace:** is there a subspace of  $\mathcal{H} = \mathcal{H}_n \otimes \mathcal{H}_{\text{el}}$  which is invariant under the dynamics up to errors  $\varepsilon^N$ , for  $N > 1$ ?
- (ii) **Intra-band dynamics:** in the affirmative case, is there any simple and convenient way to accurately describe the dynamics inside this subspace?

As for the first question, one may show that to any globally **isolated** energy band  $E_n$  corresponds a subspace of the Hilbert space which is almost-invariant under the dynamics. More precisely, one

constructs an orthogonal projector  $\Pi_{n,\varepsilon} \in \mathcal{B}(\mathcal{H})$ , with  $\Pi_{n,\varepsilon} = P_n + O(\varepsilon)$ , such that for any  $N \in \mathbb{N}$  there exists  $C_N$  such that

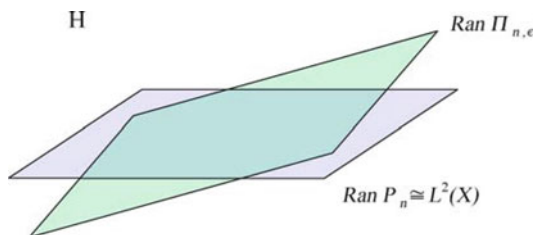
$$\begin{aligned} & \| (1 - \Pi_{n,\varepsilon}) e^{-1H_\varepsilon t/\varepsilon} \Pi_{n,\varepsilon} \Psi_0 \| \\ & \leq C_N \varepsilon^N (1 + |t|) (1 + \mathcal{E}) \| \Psi_0 \|. \end{aligned} \quad (15)$$

Here  $\mathcal{E}$  denotes a cut-off on the kinetic energy of the nuclei, which corresponds to the physical assumption that the kinetic energies of nuclei and electrons are comparable. Equation (15) shows that if the molecule is initially in a state  $\Psi_0 \in \text{Ran } \Pi_{n,\varepsilon}$ , then after a macroscopic time  $t$  the molecule is in a state which is still in  $\text{Ran } \Pi_{n,\varepsilon}$  up to an error smaller than any power of  $\varepsilon$ , with the error scaling linearly with respect to time and to the kinetic energy cut-off. For this reason the space  $\text{Ran } \Pi_{n,\varepsilon}$  is called **super-adiabatic subspace** or **almost-invariant subspace**.

We emphasize that the adiabatic decoupling, as formulated in Eq. (15), holds on a long time-scale, as opposed to the semiclassical limit which is known to hold on a time scale of order  $O(\ln \varepsilon)$ . Indeed the adiabatic decoupling is a pure quantum-phenomenon, conceptually and mathematically independent from the semiclassical limit (Fig. 3).

The previous result is based on a long history of mathematical research, starting with pioneering ideas of Sjöstrand (Brummelhuis and Nourrigat 1999; Emmrich and Weinstein 1996; Littlejohn and Flynn 1991; Martinez and Sordoni 2002; Nenciu and Sordoni 2004; Sjöstrand 1993; Sordoni 2003). It has been formulated in the form above in (Panati et al. 2003).

As for question (ii), one has to face the problem that there is no natural identification between the super-adiabatic subspace  $\text{Ran } \Pi_{n,\varepsilon}$



**Perturbation Theory and Molecular Dynamics, Fig. 3** A schematic illustration of the superadiabatic subspace  $\text{Ran } \Pi_{n,\varepsilon}$ , tilted by a correction of order  $\varepsilon$  with respect to the usual adiabatic subspace  $\text{Ran } P_n$

and  $\mathcal{H}_n \cong L^2(X)$ , and therefore no evident reduction in the number of degrees of freedom. This difficulty can be circumvented by constructing a unitary operator which intertwines the previous two spaces, namely

$$U_{n,\varepsilon} : \text{Ran } \Pi_{n,\varepsilon} \rightarrow \mathcal{H}_n \cong L^2(X). \quad (16)$$

Notice that such a unitary operator is not unique. With the help of  $U_{n,\varepsilon}$  we can map the intraband dynamics to the nuclei Hilbert space, obtaining an effective Hamiltonian  $\widehat{H}_{\text{eff},\varepsilon} := U_{n,\varepsilon} \Pi_{n,\varepsilon} H_\varepsilon \Pi_{n,\varepsilon} U_{n,\varepsilon}^{-1}$  acting in  $L^2(X)$ . It follows from Eq. (15) that for every  $N \in \mathbb{N}$  there exist  $C_N$  such that

$$\begin{aligned} & \left\| \left( e^{-iH_\varepsilon t/\varepsilon} - U_{n,\varepsilon}^{-1} e^{-i\widehat{H}_{\text{eff},\varepsilon} t/\varepsilon} U_{n,\varepsilon} \right) \Pi_{n,\varepsilon} \Psi_0 \right\|_{\mathcal{H}} \\ & \leq C_N \varepsilon^N (1+|t|)(1+\mathcal{E}) \|\Psi_0\|. \end{aligned}$$

However, with an arbitrary choice of  $U_{n,\varepsilon}$  the effective Hamiltonian  $\widehat{H}_{\text{eff},\varepsilon}$  does not appear simpler than  $\Pi_{n,\varepsilon} H_\varepsilon \Pi_{n,\varepsilon}$ . On the other side, the non-uniqueness of  $U_{n,\varepsilon}$  can be conveniently exploited to simplify the structure of the effective Hamiltonian. It has been proved in (Panati et al. 2003) that the unitary operator  $U_{n,\varepsilon}$  can be explicitly constructed in such a way that  $\widehat{H}_{\text{eff},\varepsilon}$  has a simple structure, namely it is (close to) the  $\varepsilon$ -Weyl quantization of a function

$$H_{\text{eff},\varepsilon} : X \times \mathbb{R}^{3K} \rightarrow \mathbb{R}, \quad (q, p) \mapsto H_{\text{eff},\varepsilon}(q, p),$$

defined over the classical phase space. We recall that the  $\varepsilon$ -Weyl quantization maps a (smooth) function over  $X \times \mathbb{R}^{3K}$  into a (possibly unbounded) operator acting in  $L^2(X)$ . The correspondence is such that any function  $f(q)$  is mapped into the multiplication operator times  $f(x)$ , and any function  $g(p)$  is mapped into  $g(1\varepsilon \nabla_x)$ ; for a generic function  $f(q, p)$  the ordering ambiguity is fixed by choosing

$$e^{1x \cdot q} e^{i\beta \cdot p} \mapsto e^{1(x \cdot x + \beta \cdot (1\varepsilon \nabla_x))}$$

For readers interested in the mathematical structure of Weyl quantization, we recommend (Folland 1989).

Equipped with this terminology, we come back to the effective Hamiltonian. It turns out that, with the appropriate choice of  $U_{n,\varepsilon}$ ,  $\widehat{H}_{\text{eff},\varepsilon}$  is the  $\varepsilon$ -Weyl quantization of the function

$$\begin{aligned} H_{\text{eff},\varepsilon}(q, p) &= h_0(q, p) + \varepsilon h_1(q, p) \\ &\quad + \varepsilon^2 h_2(q, p) + O(\varepsilon^3) \end{aligned} \quad (17)$$

where

$$\begin{aligned} h_0(q, p) &= \frac{1}{2} p^2 + E_n(q) + V_n(q) \\ h_1(q, p) &= -1p \cdot \chi_n(q), \nabla_q \chi_n(q) =: -p \cdot \mathcal{A}_n(q) \end{aligned} \quad (18)$$

and

$$\begin{aligned} h_2(q, p) &= \frac{1}{2} \mathcal{A}_n^2(q) + \frac{1}{2} \nabla_q \chi_n(q), (1 - P_n(q)) \\ &\quad \cdot \nabla_q \chi_n(q)_{\mathcal{H}_{\text{cl}}} \\ &\quad - p \cdot \nabla_q \chi_n(q) \\ &\quad (H_{\text{cl}}(q) - E_n(q))^{-1} (1 - P_n(q)) \\ &\quad \times p \cdot \nabla_q \chi_n(q)_{\mathcal{H}_{\text{cl}}} \end{aligned}$$

The Weyl quantization of  $h_0$  provides the leading order Born–Oppenheimer Hamiltonian Eq. (13). The term  $h_1$  has a geometric origin, involving the Berry connection  $\mathcal{A}_n(x)$ , a quantity appearing in a variety of adiabatic problems (Weigert and Littlejohn 1993); this term is responsible for the screening of magnetic fields in atoms (Yin and Mead 1994). Geometric effects in molecular systems (and more generally in adiabatic systems) are an active field of research (Faure and Zhilinskii 2000; Faure and Zhilinskii 2001; Faure and Zhilinskii 2002), see “Quantum Bifurcations” and references therein. As for the second order correction  $h_2$ , the first term completes the square  $(p - \mathcal{A}_n(x))^2$  showing that the dynamics involves a covariant derivative; the second term is known as the Born–Huang term; the last term contains the reduced resolvent (i.e. the resolvent in the orthogonal complement of  $\text{Ran } P_n$ ) and is due to the fact that the super-adiabatic subspace  $\text{Ran } \Pi_{n,\varepsilon}$  is tilted with respect to  $\text{Ran } P_n$ .

The third term in  $h_2$ , namely



$$\mathcal{M}(q, p) = \langle p \cdot \nabla \chi_n(q), (H_{\text{el}}(q) - E_n(q))^{-1} \times (1 - P_n(q)) p \cdot \nabla \chi_n(q) \rangle_{\mathcal{H}_{\text{el}}}, \quad (19)$$

appeared firstly in (Panati et al. 2003), as a consequence of the rigorous adiabatic perturbation theory developed there. This term is responsible for an  $O(\varepsilon^2)$ -correction to the effective mass of the nuclei. Indeed, since different quantization schemes differ by a term of order  $O(\varepsilon)$ , we may replace the Weyl quantization with the simpler symmetric quantization, namely we consider

$$\begin{aligned} (\widehat{M}\psi)(x) &= \sum_{\ell, k=1}^{3K} \frac{1}{2} (m_{\ell k}(x) (-i\varepsilon \partial_{x_\ell}) (-i\varepsilon \partial_{x_k}) \\ &\quad + (-i\varepsilon \partial_{x_\ell}) (-i\varepsilon \partial_{x_k}) m_{\ell k}(x)) \psi(x) \end{aligned} \quad (20)$$

where  $m$  is the  $x$ -dependent matrix

$$m_{\ell k}(x) = \partial_\ell \chi_n(x), \quad (H_e(x) - E_n(x))^{-1} \times (1 - P_n(x)) \partial_k \chi_n(x)_{\mathcal{H}_{\text{el}}}.$$

It is clear from Eq. (20) that this term induces a correction of order  $O(\varepsilon^2)$  to the Laplacean, i.e. to the inertia of the nuclei.

Finally, we point out that the effective Hamiltonian  $\widehat{H}_{\text{eff}, \varepsilon}$  can be conveniently truncated at any order in  $\varepsilon$ , getting corresponding errors in the effective quantum dynamics: if we pose

$$\widehat{H}_{\text{eff}, \varepsilon} = \sum_{j=0}^N \varepsilon^j \widehat{h}_j + O(\varepsilon^{N+1}) =: \widehat{h}_{(N), \varepsilon} + O(\varepsilon^{N+1}), \quad (21)$$

then there exist a constant  $\widetilde{C}_N$  such that

$$\begin{aligned} &\left\| \left( e^{-iH_e t/\varepsilon} - U_{n, \varepsilon}^{-1} e^{-i\widehat{h}_{(N), \varepsilon} t/\varepsilon} U_{n, \varepsilon} \right) \Pi_{n, \varepsilon} \Psi_0 \right\|_{\mathcal{H}} \\ &\leq \widetilde{C}_N \varepsilon^{N+1} (1 + |t|) (1 + \mathcal{E}) \|\Psi_0\| \end{aligned}$$

The determination of the effective Hamiltonian, here described following (Panati et al. 2003; Varandas et al. 1987), has been investigated

earlier in (Littlejohn and Flynn 1991; Weigert and Littlejohn 1993) with different but related techniques. The result in (Panati et al. 2003) is based on an iterative algorithm inspired by classical perturbation theory (see “Kolmogorov–Arnold–Moser (KAM) Theory” and “Normal Forms in Perturbation Theory”).

## Future Directions

Generically energy surfaces cross each other (see Fig. 2), and a globally isolated energy band is just a mathematical idealization. On the other side, if the initial datum  $\Psi_0(x, y) = \varphi_0(x) \chi_n(x, y)$  contains a nucleonic wavefunction  $\varphi_0$  localized far away from the crossing points, the adiabatic approximation is still valid, up to the time when the wavefunction becomes relevant in a neighborhood of radius  $\sqrt{\varepsilon}$  of the crossing points. This “hitting-time” can be estimated semiclassically, as done for example in (Spohn and Teufel 2001).

When the wavefunction reaches the region around the crossing point a relevant part of it might undergo a transition to the other crossing band. (The simultaneous crossing of more than two bands is not generic, see (von Neumann and Wigner 1929), so we focus on the crossing of two bands). The understanding of the dynamics near a conical crossing is a very active field of research.

The first step is a convenient classification of the possible structures of band crossings. Since the early days of Quantum Mechanics (von Neumann and Wigner 1929), it has been realized that eigenvalue crossings occur on submanifolds of various codimension, according to the symmetry of the problem. In the case of a molecular Hamiltonian in the form Eq. (2), generic crossings of bands with the minimal multiplicity allowed by the symmetry group have been classified in (Hagedorn 1992).

The second step consists in an analysis of the propagation of the wavefunction near the conical crossing, assuming that the initial state is concentrated on a single band, say the  $n$ th band. A pioneering work (Hagedorn 1994) shows that the qualitative picture is the following:

for crossings of codimension 1 the wavefunction follows, at the leading order, the analytic continuation of the  $n$ th band, as if there was no crossing. In the higher codimension case, a part of the wavefunction of order  $O(1)$  undergoes a transition to the other band. More recently, propagation through conical crossings has been investigated with new techniques (de Verdière 2003, 2004; de Verdière et al. 1999; Fermanian–Kammerer and Gérard 2002; Fermanian–Kammerer and Lasser 2003) opening the way to future research.

Alternatively, one may consider a family of energy bands which cross each other, but which are separated by an energy gap from the rest of the spectrum. Indeed, in a molecular collision or in excitations through a laser pulse only a few energy surfaces take part in the subsequent dynamics. Thus we take a set  $I$  of adjacent energy surfaces and call

$$P_I = \sum_{j \in I} P_j \quad (22)$$

the projection onto the relevant subspace (or subspace of physical interest). To ensure that other bands are not involved, we assume them to have a spectral gap of size  $a_{\text{gap}} > 0$  away from the energy surfaces in  $I$ , i.e.

$$\sup_{x \in X} |E_i(x) - E_j(x)| \geq a_{\text{gap}} \quad \text{for all } j \in I, i \in I^c. \quad (23)$$

Also the continuous spectrum is assumed to be at least  $a_{\text{gap}}$  away from the relevant energy surfaces. Under such assumption, the **multiband adiabatic theory** assures that the subspace  $\text{Ran } P_I$  is adiabatically protected against transitions, i.e.

$$\|(1 - P_I)e^{-iH_\varepsilon t/\varepsilon} P_I \Psi_0\|_{\mathcal{H}} = O(\varepsilon)$$

Analogously to the case of a single band, one may construct the corresponding superadiabatic projector. The effective Hamiltonian  $\hat{H}_{\text{eff},\varepsilon}$  corresponding to a family of  $m$  bands becomes, in this context, the  $\varepsilon$ -Weyl quantization of *matrix-valued* function over the classical phase space (Panati et al. 2003; Varandas et al. 1987).

A deeper understanding of nuclear dynamics near conical crossings and a further developments of multiband adiabatic perturbation theory are, in the opinion of the author, two of the main directions for future research.

## Bibliography

### Primary Literature

- Berry MV, Lim R (1990) The Born–Oppenheimer electric gauge force is repulsive near degeneracies. *J Phys A* 23: L655–L657
- Born M, Oppenheimer R (1927) Zur Quantentheorie der Molekeln. *Ann Phys (Leipzig)* 84:457–484
- Brummelhuis R, Nourrigat J (1999) Scattering amplitude for Dirac operators. *Commun Partial Differ Eq* 24(1–2):377–394
- Combes JM (1977) The Born–Oppenheimer approximation. *Acta Phys Austriaca* 17:139–159
- Combes JM, Duclos P, Seiler R (1981) The Born–Oppenheimer approximation. In: Velo G, Wightman A (eds) *Rigorous atomic and molecular physics*. Plenum, New York, pp 185–212
- de Verdière YC (2003) The level crossing problem in semiclassical analysis. I. The symmetric case. *Proceedings of the international conference in honor of Frédéric Pham (Nice, 2002)*. *Ann Inst Fourier (Grenoble)* 53(4):1023–1054
- de Verdière YC (2004) The level crossing problem in semiclassical analysis. II. The Hermitian case. *Ann Inst Fourier (Grenoble)* 54(5):1423–1441
- de Verdière YC, Lombardi M, Pollet C (1999) The microlocal Landau–Zener formula. *Ann Inst H Poincaré Phys Theor* 71:95–127
- Emmrich C, Weinstein A (1996) Geometry of the transport equation in multicomponent WKB approximations. *Commun Math Phys* 176:701–711
- Faure F, Zhilinskii BI (2000) Topological Chern indices in molecular spectra. *Phys Rev Lett* 85:960–963
- Faure F, Zhilinskii BI (2001) Topological properties of the Born–Oppenheimer approximation and implications for the exact spectrum. *Lett Math Phys* 55:219–238
- Faure F, Zhilinskii BI (2002) Topologically coupled energy bands in molecules. *Phys Lett* 302:242–252
- Fermanian–Kammerer C, Gérard P (2002) Mesures semi-classiques et croisement de modes. *Bull Soc Math France* 130:123–168
- Fermanian–Kammerer C, Lasser C (2003) Wigner measures and codimension 2 crossings. *J Math Phys* 44: 507–527
- Folland GB (1989) *Harmonic analysis in phase space*. Princeton University Press, Princeton
- Hagedorn GA (1980) A time dependent Born–Oppenheimer approximation. *Commun Math Phys* 77:1–19
- Hagedorn GA (1986) High order corrections to the time-dependent Born–Oppenheimer approximation. I. Smooth potentials. *Ann Math (2)* 124(3):571–590

- Hagedorn GA (1987) High order corrections to the time-independent Born–Oppenheimer approximation I: Smooth potentials. *Ann Inst H Poincaré Sect A* 47:1–16
- Hagedorn GA (1988a) High order corrections to the time-independent Born–Oppenheimer approximation II: diatomic Coulomb systems. *Commun Math Phys* 116: 23–44
- Hagedorn GA (1988b) High order corrections to the time-dependent Born–Oppenheimer approximation. II Coulomb systems. *Commun Math Phys* 117(3):387–403
- Hagedorn GA (1989) Adiabatic expansions near eigenvalue crossings. *Ann Phys* 196:278–295
- Hagedorn GA (1992) Classification and normal forms for quantum mechanical eigenvalue crossings. *Méthodes semi-classiques*, vol 2 (Nantes, 1991). *Astérisque* 210(7):115–134
- Hagedorn GA (1994) Molecular propagation through electron energy level crossings. *Memoirs Am Math Soc* 111:1–130
- Hagedorn GA, Joye A (2007) Mathematical analysis of Born–Oppenheimer approximations. Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon’s 60th birthday. In: *Proceedings of symposium in pure mathematics 76, Part 1*. American Mathematical Society, Providence, pp 203–226
- Herrin J, Howland JS (1997) The Born–Oppenheimer approximation: straight-up and with a twist. *Rev Math Phys* 9:467–488
- Kato T (1950) On the adiabatic theorem of quantum mechanics. *Phys Soc Jpn* 5:435–439
- Klein M, Martinez A, Seiler R, Wang XP (1992) On the Born–Oppenheimer expansion for polyatomic molecules. *Commun Math Phys* 143:607–639
- Klein M, Martinez A, Wang XP (1993) On the Born–Oppenheimer approximation of wave operators in molecular scattering theory. *Commun Math Phys* 152:73–95
- Klein M, Martinez A, Wang XP (1997) On the Born–Oppenheimer approximation of diatomic wave operators II. Singular potentials. *J Math Phys* 38:1373–1396
- Lasser C, Teufel S (2005) Propagation through conical crossings: an asymptotic transport equation and numerical experiments. *Commun Pure Appl Math* 58:1188–1230
- Littlejohn RG, Flynn WG (1991) Geometric phases in the asymptotic theory of coupled wave equations. *Phys Rev* 44:5239–5255
- Martinez A, Sordani V (2002) A general reduction scheme for the time-dependent Born–Oppenheimer approximation. *Comptes Rendus Acad Sci Paris* 334:185–188
- Mead CA, Truhlar DG (1979) On the determination of Born–Oppenheimer nuclear motion wave functions including complications due to conical intersections and identical nuclei. *J Chem Phys* 70:2284–2296
- Nenciu G, Sordani V (2004) Semiclassical limit for multistate Klein–Gordon systems: almost invariant subspaces and scattering theory. *J Math Phys* 45: 3676–3696
- Panati G, Spohn H, Teufel S (2003) Space-adiabatic perturbation theory. *Adv Theor Math Phys* 7:145–204
- Sjöstrand J (1993) Projecteurs adiabatiques du point de vue pseudodifférentiel. *Comptes Rendus Acad Sci Paris, Série I* 317:217–220
- Sordani V (2003) Reduction scheme for semiclassical operator-valued Schrödinger type equation and application to scattering. *Commun Partial Differ Eq* 28(7–8):1221–1236
- Spohn H, Teufel S (2001) Adiabatic decoupling and time-dependent Born–Oppenheimer theory. *Commun Math Phys* 224:113–132
- Varandas AJC, Brown FB, Mead CA, Truhlar DG, Blais NC (1987) A double many-body expansion of the two lowest-energy potential surfaces and nonadiabatic coupling for  $H_3$ . *J Chem Phys* 86:6258–6269
- von Neumann J, Wigner EP (1929) Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. *Phys Z* 30: 467–470
- Weigert S, Littlejohn RG (1993) Diagonalization of multi-component wave equations with a Born–Oppenheimer example. *Phys Rev A* 47:3506–3512
- Yin L, Mead CA (1994) Magnetic screening of nuclei by electrons as an effect of geometric vector potential. *J Chem Phys* 100:8125–8131

## Books and Reviews

- Bohm A, Mostafazadeh A, Koizumi A, Niu Q, Zwanziger J (2003) *The geometric phase in quantum systems. Texts and monographs in physics*. Springer, Heidelberg
- Teufel S (2003) *Adiabatic perturbation theory in quantum dynamics. Lecture notes in mathematics*, vol 1821. Springer, Berlin



# Quantum Adiabatic Theorem

Stefan Teufel

Fachbereich Mathematik, Universität Tübingen,  
Tübingen, Germany

## Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Kato's Quantum Adiabatic Theorem

Berry's Connection and Parallel Transport

Super-Adiabatic Expansions

Generalizations and Further Aspects

Future Directions

Bibliography

## Glossary

**Adiabatic decoupling** The fact that dynamical transitions between spectral subspaces of a time-dependent Hamiltonian that satisfy the gap condition are suppressed in the adiabatic limit.

**Adiabatic limit** The asymptotic regime in which the Hamiltonian of a system changes slowly compared to the relevant internal time scales. Mathematically, this is implemented by considering the asymptotic regime of small adiabatic parameter.

**Adiabatic parameter** The small parameter in the time-dependent Schrödinger equation that controls the separation of the time scales and is denoted by  $\varepsilon$  in the following text.

**Gap condition** The assumption that an eigenvalue or, more generally, a closed subset of the spectrum of a Hamiltonian has a positive distance from the rest of the spectrum.

**Hamiltonian** The generator of the time-evolution appearing in the Schrödinger equation. It is a possibly unbounded self-adjoint

operator on a Hilbert space and is usually denoted by  $H$ .

**Parallel transport** In a vector bundle with connection, there is a unique geometric way to move vectors along smooth curves in the base manifold, called the parallel transport.

**Spectral projection** Each subset of the spectrum of a self-adjoint operator is associated with an orthogonal projection, called the spectral projection onto that part of the spectrum. For an eigenvalue, the spectral projection is the orthogonal projection onto the corresponding eigenspace.

## Definition of the Subject and Its Importance

Separation of scales plays a fundamental role in understanding the dynamical behavior of complex systems in physics and other natural sciences. Adiabatic theory is primarily concerned with the separation of time scales. More specifically, it is concerned with systems in which some degrees of freedom or external parameters change on a time scale much longer than the typical frequencies in the system. Therefore, we often speak of slow and fast degrees of freedom. The goal of adiabatic theory is to find simpler effective laws for the dynamics of the slow degrees of freedom that approximate the true dynamics when the time scales are sufficiently well separated.

The quantum adiabatic theorem in its standard form refers to quantum systems with time-dependent Hamiltonians. More precisely, suppose that a smooth family of self-adjoint operators  $t \mapsto H(t)$  has a smooth family of spectral projections  $t \mapsto P(t)$  satisfying a so-called gap condition. Then, any initial state  $\psi(t_0) \in \text{ran } P(t_0)$  evolves under the Schrödinger dynamics Eq. (2) into a state  $\psi(t)$  that is localized in the subspace  $\text{ran } P(t)$  up to terms that are small in the adiabatic regime  $\varepsilon \ll 1$ , i.e., the regime of slow variation of  $H(t)$ .

Since the pioneering works of Born and Fock (1928) and Kato (1950), quantum adiabatic theory

has become an important and general framework concerned with the construction of almost invariant or almost equivariant subspaces for quantum dynamics generated by Hamiltonians that vary slowly in time or space and satisfy global or local spectral conditions such as the gap condition. The adiabatic theorem and its generalizations have crucial applications in various areas of physics. Examples include the Born-Oppenheimer approximation in molecular dynamics, effective band models and the tight-binding approximation in solid-state physics, effective models for quantum systems with constraints, adiabatic quantum computation, the Gell-Mann and Low formulas, and linear response theory for gapped ground states at zero temperature, to name just a few prominent ones.

Last but not least, the close relation of adiabatic dynamics to the geometry of vector bundles, expressed in terms such as Berry connection and Berry phase, is not only a beautiful piece of abstract mathematics but also has profound consequences for physics.

## Introduction

We briefly sketch the standard adiabatic theorem for quantum systems found in most physics textbooks. Consider a quantum system with Hilbert space  $\mathcal{H}$  and smooth one-parameter family of self-adjoint Hamiltonians  $t \mapsto H(t) \in \mathcal{L}(\mathcal{H})$ ,  $t \in I$  and  $I \subset \mathbb{R}$  an interval. One is interested in solutions of the time-dependent Schrödinger equation, i.e., the initial value problem

$$i \frac{d}{ds} \tilde{\psi}(s) = H(\varepsilon s) \tilde{\psi}(s), \quad \tilde{\psi}(s_0) = \psi_0, \quad (1)$$

for  $s \in I/\varepsilon$ , where the adiabatic parameter  $0 < \varepsilon \ll 1$  controls the time scale on which the Hamiltonian changes relative to certain internal frequencies. The smaller  $\varepsilon$  is, the longer is the time interval  $I/\varepsilon$  in which the variation of the Hamiltonian takes place. It is convenient to make a change of variables and switch to the so-called slow time scale  $t := \varepsilon s$ . Then  $\tilde{\psi}$  solves Eq. (1) if and only if  $\psi(t) := \tilde{\psi}(t/\varepsilon)$  solves

$$i\varepsilon \frac{d}{dt} \psi(t) = H(t) \psi(t), \quad \psi(t_0) = \psi_0, \quad (2)$$

for  $t \in I$ . On the slow time scale, the time interval of interest  $I$  is fixed and the adiabatic parameter  $\varepsilon$  enters the problem exclusively as a prefactor of the time derivative in Eq. (2).

The standard adiabatic theorem states that gapped spectral subspaces of  $H(t)$  are almost equivariant under the dynamics Eq. (2) whenever  $\varepsilon \ll 1$ . In the simplest case of an isolated eigenvalue, the *gap condition* takes the following form. One assumes that for each  $t \in I$ , the operator  $H(t)$  has an eigenvalue  $E(t)$  such that  $t \mapsto E(t)$  is continuous and satisfies

$$\inf_{t \in I} \text{dist}(E(t), \sigma(H(t)) \setminus \{E(t)\}) \geq g > 0. \quad (3)$$

Then, at least heuristically, the relevant internal frequency is  $g/\hbar$  and one expects the adiabatic approximation to be valid for  $\varepsilon \ll g/\hbar$ . While it turns out that this quantitative criterion for adiabatic behavior is too naive, the following statements hold under the above conditions. Let  $P(t) := \chi_{\{E(t)\}}(H(t))$  denote the spectral projection of  $H(t)$  onto the eigenspace of  $E(t)$ . Then there exists a constant  $C$  such that for any  $\psi_0 \in \text{ran } P(t_0)$  the solution of Eq. (2) satisfies

$$\|(1 - P(t))\psi(t)\| \leq \varepsilon C \|\psi_0\| \times (1 + |t - t_0|) \quad (4)$$

i.e., if the initial datum  $\psi_0$  is an eigenstate for  $H(t_0)$  with energy  $E(t_0)$ , then the solution to Eq. (2) at time  $t$  remains approximately an eigenstate of  $H(t)$  with energy  $E(t)$  whenever  $\varepsilon$  is sufficiently small. Put differently, gapped eigenspaces of  $H(t)$  are approximately equivariant under adiabatic quantum dynamics. The square of the left-hand side of Eq. (4) is sometimes called the probability of a nonadiabatic transition. There are explicit upper bounds on the constant  $C$  in Eq. (4), but they are not very accurate in most cases, and, as will be discussed below, even the order of the error in Eq. (4) as a function of  $\varepsilon$  is far from optimal in most applications.

Since the work of Kato (1950) and its geometric interpretation by Berry (1984) and Simon (1983), it

is known that the evolution of  $\psi(t)$  within the subspace  $P(t)$  is of geometric nature, more precisely that it follows the parallel transport within a vector bundle defined by  $P(t)$ . We will discuss Kato's formulation of the adiabatic theorem and its relation to parallel transport and Berry phase in sections “[Kato's Quantum Adiabatic Theorem](#)” and “[Berry's Connection and Parallel Transport](#).”

Afterwards we discuss super-adiabatic theory, that is adiabatic approximations to higher order in the adiabatic parameter  $\varepsilon$ . In contrast to Kato's approach, super-adiabatic expansions reveal also the close connection of adiabatic perturbation theory to standard regular perturbation theory.

Finally, we collect a number of more recent developments concerning generalizations and applications of adiabatic perturbation theory, including space-adiabatic perturbation theory, adiabatic theorems without a spectral gap assumption, adiabatic theorems for resonances, and adiabatic theorems for extended many-body quantum systems.

## Kato's Quantum Adiabatic Theorem

In this section, we present a version of Kato's adiabatic theorem (Kato 1950) that contains a later generalization due to Nenciu (1980) and that naturally anticipates the geometric content of the adiabatic theorem which became important and popular in physics under the name Berry phase, cf. section “[Berry's Connection and Parallel Transport](#).”

For simplicity, we assume that we are given a quantum Hamiltonian operator depending smoothly on time, i.e., a map  $H : I \rightarrow \mathcal{L}(\mathcal{H})$ ,  $t \mapsto H(t)$ , from a closed interval  $I \subset \mathbb{R}$  into the self-adjoint bounded operators on the Hilbert space  $\mathcal{H}$  that is twice differentiable. We are interested in the unitary evolution family generated by  $H(t)$ , i.e., in the solution of

$$i\varepsilon \frac{d}{dt} U^\varepsilon(t, t_0) = H(t) U^\varepsilon(t, t_0),$$

with  $U^\varepsilon(t_0, t_0) = \mathbf{1}_{\mathcal{H}}$  and  $\varepsilon > 0$ .

For any  $\psi_0 \in \mathcal{H}$ ,  $\psi(t) := U^\varepsilon(t, t_0)\psi_0$  solves the Schrödinger equation (2).

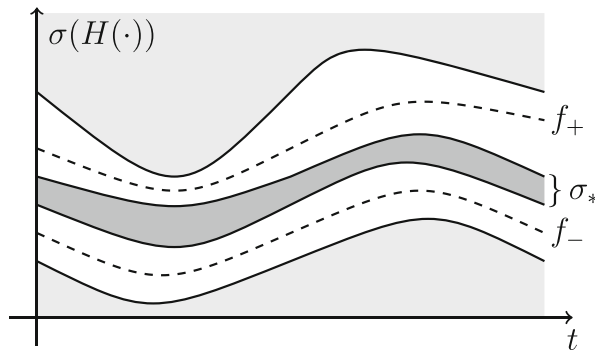
The gap condition (3) for an isolated eigenvalue can be generalized to an arbitrary spectral island  $\sigma_*(t) \subset \sigma(H(t))$  that satisfies the **gap condition**: assume that there exist continuous functions  $f_\pm : I \rightarrow \mathbb{R}$  such that for all  $t \in I$ , the functions  $f_\pm$  separate  $\sigma_*(t)$  from the rest of the spectrum, cf. Figure 1, in the sense that

$$f_\pm(t) \notin \sigma(H(t)),$$

$$\sigma_*(t) \subset [f_-(t), f_+(t)],$$

$$[f_-(t), f_+(t)] \cap (\sigma(H(t)) \setminus \sigma_*(t)) = \emptyset$$

Differentiability of the Hamiltonian  $H(t)$  together with the gap condition implies that also the family of spectral projections



**Quantum Adiabatic Theorem, Fig. 1** Sketch of the gap condition formulated in the text: The spectrum  $\sigma(H(t))$  has part  $\sigma_*(t) \subset \sigma(H(t))$  that is locally in  $t$  isolated by a gap from

its complement  $\sigma(H(t)) \setminus \sigma_*(t)$ . The dashed lines indicate the continuous functions  $f_+$  and  $f_-$  that do not intersect  $\sigma(H(t))$  and enclose only  $\sigma_*(t)$



$$P : I \rightarrow \mathcal{L}(\mathcal{H}), \quad t \mapsto P(t) := \chi_{\sigma_*(t)}(H(t))$$

is twice differentiable. This follows directly from the Riesz formula: for some  $t \in I$ , let  $\gamma : [0, 1] \rightarrow \mathbb{C} \setminus \sigma(H(t))$  be a closed differentiable path encircling  $\sigma_*(t)$  once in the positive sense. Then there is a neighborhood  $J \subset I$  of  $t$  such that for  $s \in J$  the path  $\gamma$  still does not intersect the spectrum of  $H(s)$  and hence

$$P(s) = \frac{i}{2\pi} \oint_{\gamma} dz (H(s) - z)^{-1}. \quad (5)$$

But then the differentiability of the resolvent,  $\frac{d}{dt}(H(t) - z)^{-1} = -(H(t) - z)^{-1} \dot{H}(t) (H(t) - z)^{-1}$ , implies the differentiability of  $P$ . Here and in the following we use the notation  $\dot{H}(t) := \frac{d}{dt}H(t)$  for time derivatives.

Thus, the **adiabatic Hamiltonian**

$$H_a(t) := H(t) + i\varepsilon [\dot{P}(t), P(t)]$$

is well defined and the associated **adiabatic evolution** family as well,

$$\begin{aligned} i\varepsilon \frac{d}{dt} U_a^\varepsilon(t, t_0) &= H_a(t) U_a^\varepsilon(t, t_0), \\ \text{with } U_a^\varepsilon(t_0, t_0) &= \mathbf{1}_{\mathcal{H}}. \end{aligned}$$

Under the above conditions, Kato proved the following two assertions that together imply the standard adiabatic theorem:

- (I) The **adiabatic evolution exactly intertwines the spectral subspaces**  $P_*(t)$  and their orthogonal complements at different times,

$$\begin{aligned} U_a^\varepsilon(t, t_0) P(t_0) &= P(t) U_a^\varepsilon(t, t_0) \\ \text{for all } t, t_0 \in I \text{ and } \varepsilon > 0. \end{aligned} \quad (6)$$

- (II) The **adiabatic evolution remains close to the physical evolution** if  $\varepsilon$  is small, i.e., there exists a constant  $C < \infty$  such that for all  $t, t_0 \in I$  and  $\varepsilon > 0$

$$\|U^\varepsilon(t, t_0) - U_a^\varepsilon(t, t_0)\| \leq C \varepsilon (1 + |t - t_0|). \quad (7)$$

Note that Eqs. (6) and (7) indeed imply the following generalization of Eq. (4): let  $\psi_0 \in \text{ran } P(t_0)$  then the solution  $\psi(t)$  of Eq. (2) satisfies Eq. (4):

$$\begin{aligned} \|(1 - P(t))\psi(t)\| &= \|(1 - P(t))U(t, t_0)P(t_0)\psi_0\| \\ &\stackrel{(7)}{\leq} \|(1 - P(t))U_a(t, t_0)P(t_0)\psi_0\| \\ &\quad + C \varepsilon (1 + |t - t_0|) \|\psi_0\| \\ &\stackrel{(6)}{=} C \varepsilon (1 + |t - t_0|) \|\psi_0\|. \end{aligned}$$

The proof of (I) is straightforward: Both sides of the equality (6) clearly agree at time  $t = t_0$ , and acting with  $i\varepsilon \frac{d}{dt}$  on both sides of Eq. (6) shows that their time-derivatives agree as well. The geometric explanation of Eq. (6) lies in the observation that the operator  $[P(t), \dot{P}(t)]$  is the generator of parallel transport within the Hermitian vector bundle defined by  $t \mapsto P(t)$ . We will discuss this in more detail in section “[Berry’s Connection and Parallel Transport](#).”

The proof of (II) starts with an application of the fundamental theorem of calculus,

$$\begin{aligned} U_a^\varepsilon(t, t_0) - U^\varepsilon(t, t_0) &= U^\varepsilon(t, t_0) \int_{t_0}^t ds \frac{d}{ds} (U^\varepsilon(t_0, s) U_a^\varepsilon(s, t_0)) \\ &= \frac{i}{\varepsilon} U^\varepsilon(t, t_0) \int_{t_0}^t ds U^\varepsilon(t_0, s) (H(s) - H_a(s)) U_a^\varepsilon(s, t_0) \\ &= U^\varepsilon(t, t_0) \int_{t_0}^t ds U^\varepsilon(t_0, s) [\dot{P}(s), P(s)] U_a^\varepsilon(s, t_0). \end{aligned} \quad (8)$$

The nontrivial part of the proof, which requires the gap assumption, is to show that Eq. (8) is indeed of order  $\varepsilon$ . The strategy is to write the integrand as a time-derivative of a small function and apply integration by parts. We refer, e.g., to Teufel (2003) for a detailed and pedagogical presentation of the proof. Let us mention here only that the key consequence of the gap condition is that the equation

$$\mathcal{L}_{H(t)}(F(t)) := [H(t), F(t)] = [\dot{P}(t), P(t)] \quad (9)$$

has a unique solution  $F(t) = \mathfrak{L}_{H(t)}^{-1}([\dot{P}(t), P(t)])$  for all  $t \in I$ . Indeed, the gap condition implies that the Liouvillian  $\mathfrak{L}_{H(t)}$  is continuously invertible on the subspace

$$\mathcal{A}_{\text{OD}}(t) := \left\{ A \in \mathcal{L}(\mathcal{H}) \mid A = P(t)AP(t)^\perp + P(t)^\perp AP(t) \right\}$$

of off-diagonal operators. And the right-hand side of Eq. (9) is clearly off-diagonal, since  $[B, P(t)] \in \mathcal{A}_{\text{OD}}(t)$  for any bounded linear operator  $B$ .

In the special case that  $\sigma_*(t) = \{E(t)\}$  is an isolated eigenvalue, Kato's proof of Eq. (7) yields also explicit bounds on the constant  $C$ : Let

$$g(t) := \text{dist}(E(t), \sigma(H(t)) \setminus \{E(t)\})$$

be the size of the gap at time  $t$ , then Eq. (7) holds with

$$C = 2 \left( \frac{\|\dot{P}(t)\|}{g(t)} + \frac{\|\dot{P}(t_0)\|}{g(t_0)} + \left| \int_{t_0}^t ds \left( \frac{2\|\dot{P}(s)\|^2}{g(s)} + \frac{\|\ddot{P}(s)\|}{g(s)} + \frac{\|\dot{P}(s)\|\|\dot{H}(s)\|}{g(s)^2} \right) \right| \right). \quad (10)$$

Such bounds are shown and discussed for example in Teufel (2003) and Jansen et al. (2007). However, in many applications, super-adiabatic approximations yield much better error estimates on nonadiabatic transition probabilities, cf. Eq. (15) in section “[Super-Adiabatic Expansions](#).”

Kato's formulation and proof of the adiabatic theorem not only laid the foundations of the geometric interpretation to be discussed in the next section, but initiated numerous further developments and generalizations of the adiabatic theorem. In particular, all statements of this section also hold for unbounded self-adjoint operators (Nenciu 1980), and the proof strategy was also used to construct so-called higher order quantum

adiabatic invariants (Avron et al. 1987; Garrido 1964; Lenard 1959) and adiabatic theorems without a spectral gap condition (Avron and Elgart 1999; Teufel 2001), see sections “[Super-Adiabatic Expansions](#)” and “[Generalizations and Further Aspects](#),” respectively.

## Berry's Connection and Parallel Transport

In this section, we briefly formulate the geometric interpretation of Kato's adiabatic evolution as first given by Simon (1983) following a work of Berry (1984).

To this end, it is instructive to start with a Hamiltonian  $H : M \rightarrow \mathcal{L}(\mathcal{H})$  that depends smoothly on a parameter  $x \in M$  which lives in a possibly higher dimensional space  $M \subseteq \mathbb{R}^n$ . Then for any smooth curve  $\gamma : I \rightarrow M$ , a time-dependent Hamiltonian is obtained by setting  $H_\gamma(t) := H(\gamma(t))$ .

If one assumes the gap condition to hold for all  $x \in M$ , then the family  $x \mapsto \text{ran} P(x)$  of subspaces of  $\mathcal{H}$  defines a smooth Hermitian vector bundle over  $M$ . More precisely, consider the trivial Hermitian bundle  $\Xi_{\mathcal{H}} := M \times \mathcal{H}$ , then the sub-bundle

$$\Xi_P := \{(x, v) \in I \times \mathcal{H} \mid v \in \text{ran} P(x)\} \subseteq \Xi_{\mathcal{H}}$$

inherits the Hermitian structure. Moreover, the trivial connection  $\nabla$  on  $\Xi_{\mathcal{H}}$  induces a connection  $\nabla^{\text{Be}}$  on  $\Xi_P$  the so-called Berry connection. It acts on a section  $\psi$  of  $\Xi_P$  i.e., on a smooth function  $\psi : M \rightarrow \mathcal{H}$  with  $\psi(x) \in \text{ran} P(x)$  for all  $x \in M$ , as

$$(\nabla^{\text{Be}} \psi)(x) := P(x)(\nabla \psi)(x).$$

In the case of  $\Xi_P$  being a line bundle, the holonomy of the Berry connection is called the Berry phase. For general vector bundles, one sometimes speaks of a non-abelian Berry phase.

The connection to adiabatic theory is now the following. For any smooth curve  $\gamma : I \rightarrow M$  in the parameter space  $M$ , one can define the

corresponding time-dependent Hamiltonian  $H_\gamma(t) := H(\gamma(t))$  and analogously  $P_\gamma(t)$  and  $E_\gamma(t)$ . It is straightforward to check that the parallel transport  $T_\gamma(t, t_0)$  along  $\gamma$  with respect to the Berry connection is generated by the operator  $[\dot{P}_\gamma(t), P_\gamma(t)]$ , i.e., that

$$\begin{aligned} \frac{d}{dt} T_\gamma(t, t_0) &= [\dot{P}_\gamma(t), P_\gamma(t)] T_\gamma(t, t_0) \\ \text{with } T_\gamma(t_0, t_0) &= \mathbf{1}_{\mathcal{H}}. \end{aligned} \quad (11)$$

Since the argument is instructive and at the heart of quantum adiabatic theory, we briefly discuss it. The key observation is that the derivative of any projection valued function is off-diagonal with respect to the block-decomposition it induces, i.e.,

$$\dot{P}(t) = \frac{d}{dt} P(t)^2 = \dot{P}(t)P(t) + P(t)\dot{P}(t)$$

immediately implies  $P(t)\dot{P}(t)P(t) = P(t)^\perp \dot{P}(t)P(t)^\perp = 0$  and, as a consequence,  $[P(t), [\dot{P}(t), P(t)]] = -\dot{P}(t)$ . Hence, the solution  $T_\gamma(t, t_0)$  of Eq. (11) satisfies

$$\begin{aligned} \frac{d}{dt} (T_\gamma(t_0, t) P_\gamma(t) T_\gamma(t, t_0)) \\ = T_\gamma(t_0, t) [P_\gamma(t), [\dot{P}_\gamma(t), P_\gamma(t)]] T_\gamma(t, t_0) \\ + T_\gamma(t_0, t) \dot{P}_\gamma(t) T_\gamma(t, t_0) = 0 \end{aligned}$$

and we conclude that  $T_\gamma(t, t_0) P_\gamma(t_0) = P_\gamma(t) T_\gamma(t, t_0)$ .

To see that  $T_\gamma(t, t_0)$  is indeed the desired parallel transport, observe that for  $\psi_0 \in P_\gamma(t_0)$  the section  $\psi(t) := T_\gamma(t, t_0) \psi_0$  of  $\Xi_{\mathcal{H}}$  along  $\gamma$  is indeed a parallel section of  $\Xi_P$  along  $\gamma$ , since

$$\begin{aligned} D_t^{\text{Ber}} \psi(t) &= D_t^{\text{Ber}} T_\gamma(t, t_0) \psi_0 = P_\gamma(t) \frac{d}{dt} T_\gamma(t, t_0) \psi_0 \\ &= P_\gamma(t) [\dot{P}_\gamma(t), P_\gamma(t)] T_\gamma(t, t_0) \psi_0 \\ &= P_\gamma(t) \dot{P}_\gamma(t) P_\gamma(t)^\perp T_\gamma(t, t_0) \psi_0 \\ &= 0. \end{aligned}$$

This together with the fact that  $[H_\gamma(t), P_\gamma(t)] = 0$  for all  $t \in I$  explains why Eq. (6) holds: Kato's adiabatic evolution restricted to  $P_\gamma(t_0)$  is just

parallel transport in the bundle  $\Xi_P$  along  $\gamma$  coupled with an internal evolution generated by  $H_\gamma(t)P_\gamma(t)$ . In the special case when  $P_\gamma(t)$  is the spectral projection onto an eigenspace of  $H_\gamma(t)$  with eigenvalue  $E_\gamma(t)$ , the adiabatic evolution is thus

$$U_a^\varepsilon(t, t_0) = e^{-\frac{i}{\varepsilon} \int_{t_0}^t E_\gamma(s) ds} T_\gamma(t, t_0), \quad (12)$$

i.e., parallel transport together with a dynamical phase. If  $\gamma : [t_0, t_1] \rightarrow M$  is a closed curve in  $M$ , then  $T_\gamma(t_1, t_0)$  is an element of the holonomy group  $\text{Hol}_{\gamma(t_0)}(\nabla^{\text{Be}})$  at the point  $\gamma(t_0)$ . Up to the dynamical phase factor  $e^{-\frac{i}{\varepsilon} \int_{t_0}^t E_\gamma(s) ds}$ , which is often neglected as being trivial, for closed loops in parameter space, the adiabatic evolution  $U_a^\varepsilon(t_1, t_0)$  therefore agrees with an  $\varepsilon$ -independent element of  $\text{Hol}_{\gamma(t_0)}(\nabla^{\text{Be}})$ . This observation and its generalizations to the space-adiabatic setting has profound physical consequences, some of which will be briefly discussed in section “[Generalizations and Further Aspects](#).” For an in-depth presentation of geometric concepts in adiabatic theory, see, for example, the book (Chruscinski and Jamiolkowski 2012) and for physical consequences (Bohm et al. 2013).

## Super-Adiabatic Expansions

According to Eq. (7), the adiabatic evolution  $U_a^\varepsilon(t, t_0)$  is a first order in  $\varepsilon$  approximation to the physical evolution  $U^\varepsilon(t, t_0)$ . If  $\sigma_*(t) = \{E(t)\}$  is an isolated eigenvalue, then  $U_a^\varepsilon(t, t_0)$  takes the very simple form (12). In particular, solving the dynamical highly oscillatory problem (2) is, in the adiabatic limit, reduced to solving the  $\varepsilon$ -independent parallel transport Eq. (11). Finally, if  $E(t)$  is a simple eigenvalue, the dynamical problem is reduced completely to a spectral one.

It is thus natural to ask, whether higher order adiabatic approximations with similar features are possible as well. The answer is yes, and there are several ways to formulate higher order adiabatic theory, see, for example, (Avron et al. 1987; Berry

1990; Garrido 1964; Joye and Pfister 1991; Lenard 1959; Lim and Berry 1991; Nenciu 1993; Sjöstrand 1993). We follow the approach in terms of almost equivariant subspaces introduced by Nenciu (1993) and use the terminology of super-adiabatic expansions coined by Berry (1990).

In addition to the general assumptions formulated in section “Kato’s Quantum Adiabatic Theorem,” we require the map  $t \rightarrow H(t)$  to be  $n + 1$ -times differentiable. Then there exists a differentiable family  $t \mapsto P_n^\varepsilon(t)$  of super-adiabatic projections such that the **super-adiabatic evolution**

$$i\varepsilon \frac{d}{dt} U_n^\varepsilon(t, t_0) = H_n^\varepsilon(t) U_n^\varepsilon(t, t_0),$$

$$\text{with } U_n^\varepsilon(t_0, t_0) = \mathbf{1}_{\mathcal{H}}.$$

generated by the **super-adiabatic Hamiltonian**

$$\begin{aligned} H_n^\varepsilon(t) &:= P_n^\varepsilon(t) H(t) P_n^\varepsilon(t) + P_n^\varepsilon(t)^\perp H(t) P_n^\varepsilon(t)^\perp \\ &+ i\varepsilon [P_n^\varepsilon(t), P_n^\varepsilon(t)] \end{aligned}$$

satisfies:

(I<sub>n</sub>) The **super-adiabatic evolution exactly intertwines the super-adiabatic subspaces**  $P_n(t)$  and their orthogonal complements at different times,

$$\begin{aligned} U_n^\varepsilon(t, t_0) P_n(t_0) &= P_n(t) U_n^\varepsilon(t, t_0) \\ \text{for all } t, t_0 \in I \text{ and } \varepsilon > 0. \end{aligned} \quad (13)$$

(II<sub>n</sub>) The **super-adiabatic evolution remains close to the physical evolution** if  $\varepsilon$  is small, i.e., there exists a constant  $C_n < \infty$  such that for all  $t, t_0 \in I$  and  $\varepsilon > 0$

$$\|U^\varepsilon(t, t_0) - U_n^\varepsilon(t, t_0)\| \leq C_n \varepsilon^n |t - t_0|. \quad (14)$$

It is beyond the scope of this introduction to quantum adiabatic theory to explain the construction of  $P_n^\varepsilon(t)$  in detail, but a few remarks are in order. Most importantly it should be noted that  $P_n^\varepsilon(t)$  is constructed as a perturbation series with coefficients that are **local in time**, i.e., depend

only on  $H(t)$  and its first  $n$  derivatives at time  $t$ . Moreover, if the Hamiltonian is stationary at  $t$ , i.e., all its time derivatives vanish at time  $t$ , then  $P_n^\varepsilon(t)$  agrees with the spectral projection  $P(t)$ . This implies immediately a different formulation of a higher order adiabatic theorem (cf. Avron et al. (1987) and Joye and Pfister (1991)): Assume that at times  $t_0$  and  $t_1$  the Hamiltonian is stationary, then  $P_n^\varepsilon(t_0) = P(t_0)$  and  $P_n^\varepsilon(t_1) = P(t_1)$ . Thus, Eqs. (13) and (14) imply that the physical time-evolution intertwines the spectral subspace up to terms of order  $\varepsilon^n$ ,

$$\|U^\varepsilon(t_1, t_0) P(t_0) - P(t_1) U^\varepsilon(t_1, t_0)\| \leq C_n \varepsilon^n |t_1 - t_0|. \quad (15)$$

In particular, the nonadiabatic transitions as in Eq. (4) are actually much smaller than predicted by first-order adiabatic theory.

The perturbation series for the super-adiabatic projection  $P_n^\varepsilon(t)$  is similar to the one obtained from standard Rayleigh-Schrödinger perturbation theory for isolated eigenvalues, e.g., Kato (1966), and the gap condition enters in each step of the perturbative expansion when one needs to invert the Liouvillian  $\mathcal{L}_{H(t)}$ , cf. section “Kato’s Quantum Adiabatic Theorem.”

Important examples for applications of super-adiabatic approximations in physics are the quantum Hall effect (Klein and Seiler 1990), the Gell-Mann and Low formula (Brouder et al. 2010), polarization and the piezoelectric effect (Panati et al. 2009; Schulz-Baldes and Teufel 2013), and, as to be discussed in section “Generalizations and Further Aspects,” also the adiabatic theory for extended quantum systems.

### Exponentially Small Nonadiabatic Transitions

If the map  $t \rightarrow H(t)$  is smooth or even analytic, then one can define  $P_n^\varepsilon(t)$  for any  $n \in \mathbb{N}$ . Note that in general, for fixed  $\varepsilon > 0$ , the limit  $\lim_{n \rightarrow \infty} P_n^\varepsilon(t)$  does not exist. Instead, one can optimize the error bound in Eq. (14) by choosing for every  $\varepsilon > 0$  the order  $n = n(\varepsilon)$  in such a way that  $C_n \varepsilon^n$  becomes as small as possible. The resulting  $n(\varepsilon)$  is the order of the optimal super-adiabatic projection. For analytic time dependence, one expects that  $C_n \approx n^n$

which would imply  $n(\varepsilon) \approx \frac{1}{\varepsilon^2}$  and thus exponentially small nonadiabatic transitions between super-adiabatic subspaces,  $C_n \varepsilon^n \approx e^{-1/(\varepsilon^2)}$ . Such exponential bounds on nonadiabatic transitions have been rigorously established, for example, in Joye et al. (1991), Joye and Pfister (1991, 1993), Nenciu (1993) and Sjöstrand (1993).

Actually it has been realized long before the discovery of super-adiabatic approximations that the left-hand side of Eq. (15) is exponentially small for certain simple model problems in the scattering limit  $t_0 \rightarrow -\infty$  and  $t_1 \rightarrow +\infty$ . Zener (1932) solved explicitly the Schrödinger equation generated by the Hamiltonian

$$H_{\text{LZ}} = \begin{pmatrix} g & t \\ t & -g \end{pmatrix},$$

which became known as the Landau-Zener model. He showed that the nonadiabatic transition probability in the scattering limit is  $e^{-\frac{\pi^2}{\varepsilon}}$ . The Landau-Zener Hamiltonian  $H_{\text{LZ}}$  is often used as a local model for avoided eigenvalue crossing in chemistry and physics. Rigorous proofs of Landau-Zener type formulas also for more general systems can be found, for example, in Joye (1993).

While Landau-Zener formulas give the transition probability in the scattering limit, for the simplest case of  $H(t)$  being an analytic  $2 \times 2$ -matrix valued function, the time development of exponentially small nonadiabatic transitions has been studied as well. Berry (1990) and Berry and Lim (1993) conjectured that with respect to the optimal super-adiabatic basis, the nonadiabatic transitions occur at times  $t_s$  where certain Stokes lines in the complex plane cross the real axis. Moreover, as a function of time, they have the shape of an error function, i.e., an integrated Gaussian centered at  $t_s$  and of width  $\sqrt{\varepsilon}$ . These predictions were rigorously confirmed and slightly generalized in Hagedorn and Joye (2004) and Betz and Teufel (2005).

## Generalizations and Further Aspects

### Space-Adiabatic Theorems

The mechanism of adiabatic decoupling works not only for Hamiltonians depending slowly on time,

but much more generally for Hamiltonians with “slow configurational” degrees of freedom. More precisely, let  $L^2(\mathbb{R}^n)$  be the Hilbert space for the slow configurational degrees of freedom and  $\mathcal{H}_f$  the Hilbert space of the fast degrees of freedom. Then the state space for the composite system is

$$\mathcal{H} := L^2(\mathbb{R}^n) \otimes \mathcal{H}_f = L^2(\mathbb{R}^n; \mathcal{H}_f).$$

A smooth function

$$h : \mathbb{R}^{2n} \rightarrow \mathcal{L}_{\text{sa}}(\mathcal{H})$$

defines, under appropriate additional conditions, a self-adjoint operator

$$H^\varepsilon := \text{Op}^\varepsilon(h) = “h(x, -i\varepsilon\nabla)”,$$

the  $\varepsilon$ -Weyl quantization of the symbol  $h$ . Heuristically speaking, the momenta  $-i\varepsilon\nabla$ , and thus the rate of change of the configurational degrees of freedom, are small whenever  $\varepsilon \ll 1$ .

**Example:** As the paradigmatic example for an operator appearing in physics having this structure, we briefly mention the Hamiltonian for a molecule: assuming roughly equal distribution of kinetic energies, the heavy nuclei move much slower than the light electrons. The adiabatic parameter can be chosen  $\varepsilon = \sqrt{m_e/m_n}$ , where  $m_e$  is the electron mass and  $m_n$  the typical mass of a nucleus in the molecule.

The Hilbert space for the slow nuclei is  $L^2(\mathbb{R}_x^{3N})$ ,  $N$  being the number of nuclei in the molecule, and the Hilbert space for the fast electrons is  $L^2_-(\mathbb{R}_y^{3M}; \mathbb{C}^{2M})$ , where  $M$  is the number of electrons and the subscript  $-$  indicates that because of the fermionic nature of electrons only antisymmetric wave functions are admissible. In atomic units, where  $m_e = \hbar = 1$ , the Hamiltonian for such a system of nuclei and electrons has the form

$$\begin{aligned} H_{\text{mol}}^\varepsilon = & - \sum_{j=1}^N \frac{\varepsilon^2}{2\tilde{m}_j} \Delta_{x_j} - \sum_{j=1}^M \frac{1}{2} \Delta_{y_j} \\ & + V_{\text{nn}}(x) + V_{\text{ne}}(x, y) + V_{\text{ee}}(y), \end{aligned} \quad (16)$$

where  $\tilde{m}_j := \varepsilon^2 m_j$  is of order one, and  $m_j$  denotes the mass of nucleus  $j$ . The Coulomb

interaction potentials between the different charged particles do not depend on the masses and thus not on  $\varepsilon$ . The molecular Hamiltonian

can be written the  $\varepsilon$ -Weyl quantization of the operator valued function

$$h : \mathbb{R}^{6N} \longrightarrow \mathcal{L}_{\text{sa}}\left(L^2\left(\mathbb{R}_y^{3M}; \mathbb{C}^{2M}\right)\right), (q, p) \mapsto h(q, p) := \underbrace{\sum_{j=1}^N \frac{p_j^2}{2\tilde{m}_j} - \sum_{j=1}^M \frac{1}{2} \Delta_{y_j} + V_{\text{nn}}(q) + V_{\text{ne}}(q, y) + V_{\text{ee}}(y)}_{= H_{\text{el}}(q)}.$$

The operator  $H_{\text{el}}(q)$  is the Hamiltonian describing  $M$  electrons in the static field of  $N$  nuclei with configuration  $q$ . Clearly, neither  $H_{\text{mol}}$  nor  $H_{\text{el}}(q)$  are bounded operators, a fact that makes the mathematical analysis more technical but should not distract from the structural aspects to be discussed here.

With the example of molecular dynamics in mind, we briefly sketch the main results of space-adiabatic perturbation theory when applied to Hamiltonians of the form  $H^\varepsilon = \text{Op}^\varepsilon(h)$ . The gap condition now requires that there exists a locally gapped spectral island  $\sigma_*(q, p) \subset \sigma(h(q, p))$ . Again this means that there are continuous functions  $f_\pm : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  such that for all  $(q, p) \in \mathbb{R}^{2n}$ , the functions  $f_\pm$  separate  $\sigma_*(q, p)$  from the rest of the spectrum in the sense that

$$f_\pm(q, p) \notin \sigma(h(q, p)),$$

$$\sigma_*(q, p) \subset [f_-(q, p), f_+(q, p)],$$

$$[f_-(q, p), f_+(q, p)] \cap (\sigma(h(q, p)) \setminus \sigma_*(q, p)) = \emptyset.$$

Then there exists functions  $P_n^\varepsilon(q, p) : \mathbb{R}^{2n} \rightarrow \mathcal{L}_{\text{sa}}(\mathcal{H}_f)$  such that  $\Pi_n^\varepsilon := \text{Op}^\varepsilon(P_n^\varepsilon)$  is an orthogonal projection on  $\mathcal{H} = L^2(\mathbb{R}^n, \mathcal{H}_f)$  that almost commutes with  $H^\varepsilon$ ,

$$\| [\Pi_n^\varepsilon, H^\varepsilon] \| \leq C_n \varepsilon^{n+1}. \quad (17)$$

The bound Eq. (17) implies immediately that the range of  $\Pi_n^\varepsilon$  is almost invariant under the dynamics generated by  $H^\varepsilon$ , i.e., that

$$\| [\Pi_n^\varepsilon, e^{-iH^\varepsilon t}] \| \leq C_n \varepsilon^{n+1} |t|. \quad (18)$$

Hence,  $\text{ran} \Pi_n^\varepsilon \subset \mathcal{H}$  is called an almost invariant subspace for  $H^\varepsilon$ . In summary, locally isolated spectral islands of the symbol  $h$  give rise to almost invariant subspaces for the Hamiltonian  $H^\varepsilon = \text{Op}^\varepsilon(h)$ . Defining in analogy to the adiabatic resp. super-adiabatic dynamics of sections “[Kato’s Quantum Adiabatic Theorem](#)” and “[Super-Adiabatic Expansions](#)” the effective dynamics  $e^{-iH_{\text{eff},n}^\varepsilon t}$  generated by

$$H_{\text{eff},n}^\varepsilon := \Pi_n^\varepsilon H^\varepsilon \Pi_n^\varepsilon + \Pi_n^{\varepsilon,\perp} H^\varepsilon \Pi_n^{\varepsilon,\perp},$$

one obtains similarly to Eq. (14) that

$$\| e^{-iH^\varepsilon t} - e^{-iH_{\text{eff},n}^\varepsilon t} \| \leq C_n \varepsilon^{n+1} |t|.$$

In the most relevant situation where  $\sigma_*(q, p) = \{E(q, p)\}$  is an isolated eigenvalue (or more generally a finite number of eigenvalues of finite multiplicity), one can explicitly compute the restriction of  $H_{\text{eff},n}^\varepsilon$  to  $\text{ran} \Pi_n^\varepsilon$  and thereby obtain an explicit effective Hamiltonian for the dynamics within the almost invariant subspace corresponding to  $E$ . The effective Hamiltonian within this subspace is given by a  $\varepsilon$ -Weyl operator with scalar (or matrix-valued) symbol which can be further analyzed by methods of semiclassical analysis. This construction of effective Hamiltonians based on adiabatic decoupling of local spectral subspaces has a number of important applications in physics, which we briefly list together with some references below. Before, we mention that the general scheme of space-adiabatic perturbation theory as outlined above is developed in Panati et al. (2002, 2003a), based on



Nenciu and Sordoni (2004), Martinez and Sordoni (2002) and earlier ideas in Helffer and Sjöstrand (1990), Emmrich and Weinstein (1996), and Brummelhuis and Nourrigat (1999). A general scheme including also semiclassical approximation for the slow degrees of freedom is presented in Stiepan and Teufel (2013).

In the example of molecular dynamics, the spectral island is typically an isolated electronic band  $\sigma_*(q, p) = \{E(q, p)\}$  and Eq. (17) shows that nonadiabatic transitions between electronic levels are smaller than any power of  $\varepsilon$  as long as the electronic energy levels do not intersect. The corresponding effective Hamiltonian  $H_{\text{eff},0}^\varepsilon$  to  $\text{ran}\Pi_0^\varepsilon$  is very well known as the Born-Oppenheimer approximation, and it is unitarily equivalent to

$$H_{\text{BO},0}^\varepsilon = - \sum_{j=1}^N \frac{\varepsilon^2}{2\tilde{m}_j} \Delta_{x_j} + E(x)$$

acting on  $L^2(\mathbb{R}_x^{3N})$ . Comparing with Eq. (16) shows, that in the Born-Oppenheimer approximation the electronic degrees of freedom no longer appear, but instead their influence is reduced to the effective potential  $E$ . Space-adiabatic perturbation theory now allows to compute explicitly and systematically higher order corrections to the Born-Oppenheimer approximation. For the computation of  $H_{\text{BO},2}^\varepsilon$ , we refer to Littlejohn and Weigert (1993) and more generally to Mátyus and Teufel (2019). The results in Mátyus and Teufel (2019) served as the starting point for high-precision, ab initio computations of energy levels of small molecules, see, e.g., Ferenc et al. (2020). The most general rigorous results can be found in Martinez and Sordoni (2009). For more details and standard references on adiabatic perturbation theory in molecular dynamics, see the article *Perturbation Theory and Molecular Dynamics* by Gianluca Panati. Finally, a more recent article generalizing super-adiabatic transition histories to the Born-Oppenheimer approximation is Betz and Goddard (2009).

Further examples for applications of space-adiabatic perturbation theory are systems of fermions in periodic solids (Panati et al. 2003b; Stiepan and Teufel 2013), quantum dynamics with

constraints (Wachsmuth and Teufel 2009; Lampart and Teufel 2017; Haag and Lampart 2019), and even loop-quantum gravity (Stottmeister and Thiemann 2016).

### Adiabatic Theorems Without Gap Condition

In Avron and Elgart (1999) and Bornemann (1998), it was realized independently and proved by different methods that an adiabatic theorem can also hold in the absence of a spectral gap. In the proof of the adiabatic theorem, the gap condition is used twice: to show that the differentiability of the Hamiltonian  $H(t)$  implies the differentiability of the spectral projection  $P(t)$ , see Eq. (5), and to show that the Liouvillian  $\mathcal{L}_{H(t)}$  restricted to  $\mathcal{A}_{\text{OD}}$  has a bounded inverse.

In the adiabatic theorems without gap condition mentioned above, one instead assumes that  $\sigma_*(t) = \{E(t)\}$  is an eigenvalue such that  $t \mapsto P(t)$  is twice differentiable. Moreover, although without a spectral gap the inverse of the Liouvillian is not bounded anymore, one uses a regularization of the latter and thereby pays with a worse error bound in Eq. (7), typically an error of order  $\varepsilon^\alpha$  for some  $0 < \alpha < 1$ . The precise value of  $\alpha$  depends on the details of the Hamiltonian and can be related to the Hölder-continuity of the spectral measure of  $H(t)$  near  $E(t)$ . In particular, the theorem applies if  $E(t)$  is embedded into or next to continuous spectrum of  $H(t)$  or crosses other eigenvalues of  $H(t)$ . A simplified proof of the adiabatic theorem without gap assumption with improved error bounds can be found in Teufel (2001) and a space-adiabatic version with applications to particles coupled to a massless field in Joye et al. (2020), Tenuta (2008), and Teufel (2002).

### Adiabatic Theorems for Resonances

Roughly speaking, a resonance for a Hamiltonian  $H$  is a state that is invariant under the dynamics generated by  $H$  for long times and up to small errors. Most common are resonances arising from eigenstates of a subsystem Hamiltonian that is only weakly coupled to a system with a large number of degrees of freedom or shape resonances in which a system is confined by a potential well but can escape by tunneling through the

potential barrier. Using ideas very similar to those for adiabatic theorems without spectral gap, also adiabatic theorems for resonances were established in Abou-Salem and Fröhlich (2005, 2007), and Elgart and Hagedorn (2011). It should be noted that in realistic models, at most the ground state of a Hamiltonian is really an eigenstate and all other eigenstates turn into resonances, for example, through coupling to the quantized radiation field. The latter is important in molecular dynamics, and a Born-Oppenheimer approximation for resonances, that is a space-adiabatic theorem for resonances, is developed in Teufel and Wachsmuth (2012).

### Adiabatic Theorems for Open Quantum Systems

The dynamics of an open quantum system is no longer generated by a self-adjoint operator on a Hilbert space, but the generator is merely a linear operator on a suitable Banach space, often called the Lindbladian in this context. Adiabatic theorems for the dynamics generated by suitable Lindbladians are discussed in Abou-Salem and Fröhlich (2005) and Avron et al. (2012). More generally, adiabatic theory for generators that are not self-adjoint is developed in Joye (2007) and Schmid (2019).

### Adiabatic Theorems for Many-Body Quantum Systems

More recently, adiabatic theorems designed specifically for extended many-body quantum systems with local interactions have been established. Standard adiabatic theory, providing norm estimates, breaks down in many-body systems whenever the number of degrees of freedom becomes large. This is because even tiny deviations in each factor of a tensor product with a large number of factors lead to almost orthogonal states. Put differently, the norms  $\|\dot{H}(t)\|$  and  $\|\dot{P}(t)\|$  that enter the bound (10) grow linear in the number of particles or spins in a many-body system and thus the adiabatic bounds on norm differences deteriorate in the thermodynamic limit.

In Bachmann et al. (2017, 2018), the authors use locality of the dynamics in such systems, expressed by Lieb-Robinson bounds, in order to

establish the validity of the adiabatic approximation when it comes to expectation values of local observables that act only on a fixed finite number of degrees of freedom. A super-adiabatic version of their result is obtained in Monaco and Teufel (2019) and a space-adiabatic version in Teufel (2020).

### Adiabatic Theorems for Nonlinear Dynamics

It is evident that the principle of adiabatic decoupling should also apply to certain nonlinear quantum evolutions. While to our knowledge there has not yet been developed a general nonlinear adiabatic theory, many partial results are available. Examples are weak nonlinearities (Sparber 2016), the nonlinear Landau Zener problem (Fermanian-Kammerer and Joye 2020), or the Landau-Pekar equations for the dynamics of polarons (Frank and Gang 2020; Leopold et al. 2019).

### Future Directions

The idea of decoupling the dynamics of slow and fast degrees of freedom in quantum systems is extremely general and has been applied in many different physical and mathematical contexts. However, the underlying physical and mathematical conditions and the results derived from them are very different. It is therefore not expected that an overarching mathematical adiabatic theorem can be established or is desirable that encompasses all known results. It is also difficult to foresee what new applications will require further generalizations or adaptations of adiabatic theory.

Current open problems, where progress is expected in the foreseeable future, mainly concern issues of extended many-body quantum systems and nonlinear adiabatic theory. For example, an adiabatic theorem for open extended many-body systems would be desirable to understand more realistic models of quantum Hall transport. Also, the idea of a local gap and a local adiabatic theory in such systems seems relevant, since the Hamiltonian of such a system often does not have a spectral gap due to edge states, but this should not change the adiabatic

behavior in the bulk. Finally, the influence of the range of interactions on the quality of the adiabatic approximation in many-body systems remains to be understood.

## Bibliography

- Abou-Salem W, Fröhlich J (2005) Adiabatic theorems and reversible isothermal processes. *Lett Math Phys* 72: 153–163
- Abou-Salem W, Fröhlich J (2007) Adiabatic theorems for quantum resonances. *Commun Math Phys* 273: 651–675
- Avron JE, Elgart A (1999) Adiabatic theorem without a gap condition. *Commun Math Phys* 203:445–463
- Avron J, Seiler R, Yaffe LG (1987) Adiabatic theorems and applications to the quantum Hall effect. *Commun Math Phys* 110:33–49
- Avron J, Fraas M, Graf G, Grech P (2012) Adiabatic theorems for generators of contracting evolutions. *Commun Math Phys* 314:163–191
- Bachmann S, De Roeck W, Fraas M (2017) Adiabatic theorem for quantum spin systems. *Phys Rev Lett* 119:060201
- Bachmann S, De Roeck W, Fraas M (2018) The adiabatic theorem and linear response theory for extended quantum systems. *Commun Math Phys* 361:997–1027
- Berry MV (1984) Quantal phase factors accompanying adiabatic changes. *Proc Roy Soc Lond A* 392:45
- Berry MV (1990) Histories of adiabatic quantum transitions. *Proc R Soc Lond A* 429:61–72
- Berry MV, Lim R (1993) Universal transition prefactors derived by superadiabatic renormalization. *J Phys A* 26:4737–4747
- Betz V, Goddard B (2009) Accurate prediction of non-adiabatic transitions through avoided crossings. *Phys Rev Lett* 103:213001
- Betz V, Teufel S (2005) Precise coupling terms in adiabatic quantum evolution: the generic case. *Commun Math Phys* 260:481–509
- Bohm A, Mostafazadeh A, Koizumi H, Niu Q, Zwanziger J (2013) The geometric phase in quantum systems. Springer
- Born M, Fock V (1928) Beweis des Adiabatsatzes. *Z Phys* 51:165–169
- Born M, Oppenheimer R (1927) Zur Quantentheorie der Molekeln. *Ann Phys (Leipzig)* 84:457–484
- Bornemann F (1998) Homogenization in time of singularly perturbed mechanical systems. *Lecture notes in mathematics*, vol 1687. Springer, Heidelberg
- Brouder C, Panati G, Stoltz G (2010) Gell-Mann and low formula for degenerate unperturbed states. *Ann Henri Poincaré* 10:1285–1309
- Brummelhuis R, Nourrigat J (1999) Scattering amplitude for Dirac operators. *Commun Partial Diff Equ* 24: 377–394
- Chruscinski D, Jamiolkowski A (2012) Geometric phases in classical and quantum mechanics. *Progress in mathematical physics*, vol 36. Springer, New York
- Combes J-M, Duclos P, Seiler R (1981) The Born-Oppenheimer approximation. In: Velo G, Wightman A (eds) *Rigorous atomic and molecular physics*. Plenum, New York, pp 185–212
- Elgart A, Hagedorn G (2011) An adiabatic theorem for resonances. *Commun Pure Appl Math* 64:1029–1058
- Emmrich C, Weinstein A (1996) Geometry of the transport equation in multicomponent WKB approximations. *Commun Math Phys* 176:701–711
- Ferenc D, Korobov V, Mátyus E (2020) Nonadiabatic, relativistic, and leading-order QED corrections for rovibrational intervals of  $\text{He } 4\ 2+(X\ \Sigma\ 2\ u+)$ . *Phys Rev Lett* 125:213001
- Fermanian-Kammerer C, Joye A (2020) A nonlinear quantum adiabatic approximation. *Nonlinearity* 33:4715
- Frank R, Gang Z (2020) A nonlinear adiabatic theorem for the one-dimensional Landau-Pekar equations. *J Funct Anal* 279:108631
- Garrido LM (1964) Generalized adiabatic invariance. *J Math Phys* 5:335
- Haag S, Lampart J (2019) The adiabatic limit of the connection Laplacian. *J Geom Anal* 29:2644–2673
- Hagedorn GA, Joye A (2004) Time development of exponentially small non-adiabatic transitions. *Commun Math Phys* 250:393–413
- Helffer B, Sjöstrand J (1990) Analyse semiclassique pour l'équation de Harper II. *Mém SMF* 40:148 p
- Jansen S, Ruskai M-B, Seiler R (2007) Bounds for the adiabatic approximation with applications to quantum computation. *J Math Phys* 48:102111
- Joye A (1993) Non-trivial prefactors in adiabatic transition probabilities induced by high order complex degeneracies. *J Phys A* 26:6517–6540
- Joye A (2007) General adiabatic evolution with a gap condition. *Commun Math Phys* 275:139–162
- Joye A (2021) Adiabatic Lindbladian evolution with small dissipators. *Commun Math Phys*. <https://doi.org/10.48550/arXiv.2106.15749>
- Joye A, Pfister C-E (1991) Exponentially small adiabatic invariant for the Schrödinger equation. *Commun Math Phys* 140:15–41
- Joye A, Pfister C-E (1993) Superadiabatic evolution and adiabatic transition probability between two non-degenerate levels isolated in the spectrum. *J Math Phys* 34:454–479
- Joye A, Kunz H, Pfister C-E (1991) Exponential decay and geometric aspect of transition probabilities in the adiabatic limit. *Ann Phys* 208:299
- Joye A, Merkli M, Spehner D (2020) Adiabatic transitions in a two-level system coupled to a free Boson reservoir. *Ann Henri Poincaré* 21:3157–3199
- Kato T (1950) On the adiabatic theorem of quantum mechanics. *Phys Soc Jpn* 5:435–439
- Kato T (1966) *Perturbation theory for linear operators*. Springer, Berlin/Heidelberg. (2nd edition, 1995)

- Klein M, Seiler R (1990) Power-law corrections to the Kubo formula vanish in quantum Hall systems. *Commun Math Phys* 128:141–160
- Lampart J, Teufel S (2017) The adiabatic limit of Schrödinger operators on fibre bundles. *Math Ann* 367:1647–1683
- Lenard A (1959) Adiabatic invariance to all orders. *Ann Phys* 6:261–276
- Leopold N, Rademacher S, Schlein B, Seiringer R (2019) The Landau-Pekar equations: adiabatic theorem and accuracy. *Anal PDE*. <https://doi.org/10.48550/arXiv.1904.12532>
- Lim R, Berry MV (1991) Superadiabatic tracking of quantum evolution. *J Phys A* 24:3255–3264
- Littlejohn RG, Weigert S (1993) Diagonalization of multi-component wave equations with a Born-Oppenheimer example. *Phys Rev A* 47:3506–3512
- Martinez A, Sordani V (2002) A general reduction scheme for the time-dependent Born-Oppenheimer approximation. *CR Math* 334:185–188
- Martinez A, Sordani V (2009) Twisted pseudodifferential calculus and application to the quantum evolution of molecules. *Mem AMS* 936
- Mátyus E, Teufel S (2019) Effective non-adiabatic Hamiltonians for the quantum nuclear motion over coupled electronic states. *J Chem Phys* 151:014113
- Monaco D, Teufel S (2019) Adiabatic currents for interacting fermions on a lattice. *Rev Math Phys* 31:1950009
- Nenciu G (1980) On the adiabatic theorem of quantum mechanics. *J Phys A* 13:L15–L18
- Nenciu G (1993) Linear adiabatic theory. Exponential estimates. *Commun Math Phys* 152:479–496
- Nenciu G, Sordani V (2004) Semiclassical limit for multi-state Klein-Gordon systems: almost invariant subspaces, and scattering theory. *J Math Phys* 45:3676–3696
- Panati G, Spohn H, Teufel S (2002) Space-adiabatic perturbation theory in quantum dynamics. *Phys Rev Lett* 88:250405
- Panati G, Spohn H, Teufel S (2003a) Space-adiabatic perturbation theory. *Adv Theor Math Phys* 7:145–204
- Panati G, Spohn H, Teufel S (2003b) Effective dynamics for the Bloch electron: Peierls substitution and beyond. *Commun Math Phys* 242:547–578
- Panati G, Sparber C, Teufel S (2009) Geometric currents in piezoelectricity. *Arch Rat Mech Anal* 191:387–422
- Schmid J (2019) Adiabatic theorems for general linear operators with timeindependent domains. *Rev Math Phys* 31:1950014
- Schulz-Baldes H, Teufel S (2013) Orbital polarization and magnetization for independent particles in disordered media. *Commun Math Phys* 319:649–681
- Simon B (1983) Holonomy, the quantum adiabatic theorem, and Berry's phase. *Phys Rev Lett* 51:2167–2170
- Sjöstrand J (1993) Projecteurs adiabatiques du point de vue pseudodifférentiel. *C R Acad Sci* 317:217–220
- Sparber C (2016) Weakly nonlinear time-adiabatic theory. *Ann Henri Poincaré* 17:913–936
- Stiepan H, Teufel S (2013) Semiclassical approximations for Hamiltonians with operator-valued symbols. *Commun Math Phys* 320:821–849
- Stottmeister A, Thiemann T (2016) Coherent states, quantum gravity, and the Born-Oppenheimer approximation. III: applications to loop quantum gravity. *J Math Phys* 57:083509
- Tenuta L (2008) Quasi-static limits in nonrelativistic quantum electrodynamics. *Ann Henri Poincaré* 9:553–593
- Teufel S (2001) A note on the adiabatic theorem without gap condition. *Lett Math Phys* 58:261–266
- Teufel S (2002) Effective N-body dynamics for the massless Nelson model and adiabatic decoupling without spectral gap. *Ann Henri Poincaré* 3:939–965
- Teufel S (2003) Adiabatic perturbation theory in quantum dynamics. *Lecture notes in mathematics*, vol 1821. Springer, Berlin
- Teufel S (2020) Non-equilibrium almost-stationary states and linear response for gapped quantum systems. *Commun Math Phys* 373:621–653
- Teufel S, Spohn H (2002) Semiclassical motion of dressed electrons. *Rev Math Phys* 4:1–28
- Teufel S, Wachsmuth J (2012) Spontaneous decay of resonant energy levels for molecules with moving nuclei. *Commun Math Phys* 315:699–738
- Wachsmuth J, Teufel S (2009) Effective Hamiltonians for constrained quantum systems. *Mem AMS* 1083
- Zener C (1932) Non-adiabatic crossing of energy levels. *Proc Roy Soc London* 137:696–702

## Books

- Bohm A, Mostafazadeh A, Koizumi H, Niu Q, Zwanziger J (2013) *The geometric phase in quantum systems*. Springer, Berlin/Heidelberg
- Chruscinski D, Jamiolkowski A (2012) *Geometric phases in classical and quantum mechanics. Progress in mathematical physics*, vol 36. Springer, New York
- Teufel S (2003) *Adiabatic perturbation theory in quantum dynamics. Lecture notes in mathematics*, vol 1821. Springer, Berlin

# Quantum Bifurcations

Boris Zhilinskii

Université du Littoral, Dunkerque, France

## Article Outline

Glossary

Definition of the Subject

Introduction

Simplest Effective Hamiltonians

Simplest Hamiltonians for Two Degree-of-Freedom Systems

Bifurcations and Symmetry

Bifurcation Diagrams for Two Degree-of-Freedom Integrable Systems

Bifurcations of “Quantum Bifurcation Diagrams”

Semi-Quantum Limit and Reorganization of Quantum Bands

Multiple Resonances and Quantum State Density Physical Applications and Generalizations

Future Directions

Bibliography

## Glossary

**Classical limit** The classical limit is the classical mechanical problem which can be constructed from a given quantum problem by some limiting procedure. During such a construction the classical limiting manifold should be defined which plays the role of classical phase space. As soon as quantum mechanics is more general than classical mechanics, going to the classical limit from a quantum problem is much more reasonable than discussing possible quantizations of classical theories (Simon 1980).

**Energy-momentum map** In classical mechanics for any problem which allows the existence of several integrals of motion (typically energy and other integrals often named as momenta) the Energy-Momentum (EM) map gives the correspondence between the phase space of

the initial problem and the space of values of all independent integrals of motion. The energy-momentum map introduces the natural foliation of the classical phase space into common levels of values of energy and momenta (Cushman and Bates 1997; Guillemin 1994). The image of the EM map is the region of the space of possible values of integrals of motion which includes regular and critical values. The quantum analog of the image of the energy-momentum map is the joint spectrum of mutually commuting quantum observables.

**Joint spectrum** For each quantum problem a maximal set of mutually commuting observables can be introduced (Dirac 1982). A set of quantum wave functions which are mutual eigenfunctions of all these operators exists. Each such eigenfunction is characterized by eigenvalues of all mutually commuting operators. The representation of mutual eigenvalues of  $n$  commuting operators in the  $n$ -dimensional space gives the geometrical visualization of the joint spectrum.

**Monodromy** In general, the monodromy characterizes the evolution of some object after it makes a close path around something. In classical Hamiltonian dynamics the Hamiltonian monodromy describes for completely integrable systems the evolution of the first homology group of the regular fiber of the energy-momentum map after a close path in the regular part of the base space (Cushman and Bates 1997). For a corresponding quantum problem the quantum monodromy describes the modification of the local structure of the joint spectrum after its propagation along a close path going through a regular region of the lattice.

**Quantum bifurcation** Qualitative modification of the joint spectrum of the mutually commuting observables under the variation of some external (or internal) parameters and associated in the classical limit with the classical bifurcation is named quantum bifurcation (Pavlichenkov 1993). In other words the quantum bifurcation is the manifestation of the classical bifurcation



presented in the classical dynamic system in the quantum version of the same system.

**Quantum-classical correspondence** Starting from any quantum problem the natural question consists of defining the corresponding classical limit, i.e. the classical dynamic variables forming the classical phase space and the associated symplectic structure. Whereas in simplest quantum problems defined in terms of standard position and momentum operators with commutation relation  $[q_i, p_j] = i\hbar\delta_{ij}$ ,  $[q_i, q_j] = [p_i, p_j] = 0$  ( $i, j = 1 \dots n$ ) the classical limit phase space is the  $2n$ -dimensional Euclidean space with standard symplectic structure on it, the topology of the classical limit manifold in many other important for physical applications cases can be rather non-trivial (Simon 1980; Zhang et al. 1990).

**Quantum phase transition** Qualitative modifications of the ground state of a quantum system occurring under the variation of some external parameters at zero temperature are named quantum phase transitions (Sachdev 1999). For finite particle systems the quantum phase transition can be considered as a quantum bifurcation (Pavlichenkov 2006).

**Spontaneous symmetry breaking** Qualitative modification of the system of quantum states caused by perturbation which has the same symmetry as the initial problem. Local symmetry of solutions decreases but the number of solutions increases. In the energy spectra of finite particle systems the spontaneous symmetry breaking produces an increase of the “quasidegeneracy”, i.e. formation of clusters of quasi-degenerate levels whose multiplicity can be much higher than the dimension of the irreducible representations of the global symmetry group (Michel 1980).

**Symmetry breaking** Qualitative changes in the properties (dynamical behavior, and in particular in the joint spectrum) of quantum systems which are due to modifications of the global symmetry of the problem caused by external (less symmetrical than original problem) perturbation can be described as symmetry breaking effects. Typical effects consist of splitting of degenerate energy levels classified initially according to irreducible

representation of the initial symmetry group into less degenerate groups classified according to irreducible representation of the subgroup (the symmetry group of the perturbation) (Landau and Lifschits 1981).

## Definition of the Subject

Quantum bifurcations (QB) are qualitative phenomena occurring in quantum systems under the variation of some internal or external parameters. In order to make this definition a little more precise we add the additional requirement: The qualitative modification of the “behavior” of a quantum system can be described as QB if it consists of the manifestation for the quantum system of the classical bifurcation presented in classical dynamic systems which is the classical analog of the initial quantum system. Quantum bifurcations are typical elementary steps leading from the simplest in some way effective Hamiltonian to more complicated ones under the variation of external or internal parameters. As internal parameters one may consider the values of exact or approximate integrals of motion. The construction of an effective Hamiltonian is typically based on the averaging and/or reduction procedure which results in the appearance of “good” quantum numbers (or approximate integrals of motion). The role of external parameters can be played by forces of external champs, phenomenological constants in the effective Hamiltonians, particle masses, etc. In order to limit the very broad field of qualitative changes and of possible quantum bifurcations in particular, we restrict ourselves mainly to quantum systems whose classical limit is associated with compact phase space and is nearly integrable. This means that for quantum problems the set of mutually commuting observables can be constructed within a reasonable physical approximation almost everywhere at least locally.

Quantum bifurcations are supposed to be universal phenomena which appear in generic families of quantum systems and explain how relatively simple behavior becomes complicated under the variation of some physical parameters. To know these elementary bricks responsible for

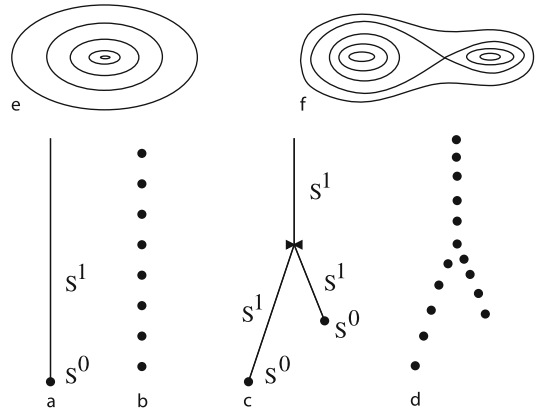


increasing complexity of quantum systems under control parameter modifications is extremely important in order to make the extrapolation to regimes inaccessible to experimental study.

## Introduction

In order to better understand the manifestations of quantum bifurcations and their significance for concrete physical systems we start with the description of several simple model physical problems which exhibit in some sense the simplest (but nevertheless) generic behavior. Let us start with the harmonic oscillator. A one-dimensional harmonic oscillator has an equidistant system of eigenvalues. All eigenvalues can be labelled by consecutive integer quantum numbers which have the natural interpretation in terms of the number of zeros of eigenfunctions. The classical limit manifold (classical phase space) is a standard Euclidean 2-dimensional space with natural variables  $\{p, q\}$ . The classical Hamiltonian for the harmonic oscillator is an example of a Morse-type function which has only one stationary point  $p = q = 0$  and all non-zero energy levels of the Hamiltonian are topological circles. If now we deform slightly the Hamiltonian in such a way that its classical phase portrait remains topologically the same, the spectrum of the quantum problem changes but it can be globally described as a regular sequence of states numbered consecutively by one integer and such description remains valid for any mass parameter value. Note, that for this problem increasing mass means increasing quantum state density and approaching classical behavior (classical limit).

More serious modification of the harmonic oscillator can lead, for example, to creation of new stationary points of the Hamiltonian. In classical theory this phenomenon is known as fold bifurcation or fold catastrophe (Arnold 1992; Gilmore 1981). The phase portrait of the classical problem changes qualitatively. As a function of energy the constant level set of the Hamiltonian has different topological structure (one circle, two circles, figure eight, circle and a point, or simply point). The quantum version of the same problem



**Quantum Bifurcations, Fig. 1** Classical and quantum bifurcations for a one degree-of-freedom system. Situations before (a, b, e) and after (c, d, f) the bifurcation are shown. (a) Energy map for harmonic oscillator-type system. Inverse images of each point are indicated. (b) Quantum state lattice for harmonic oscillator-type system. (c) Energy map after the bifurcation. Inverse images of each point are indicated. (d) Quantum state lattice after bifurcation represented as composed of three regular parts glued together. (e) Phase portrait for harmonic oscillator-type system. Inverse images are  $S^1$  (generic inverse image) and  $S^0$  (inverse image for minimal energy value). (f) Phase portrait after bifurcation

shows the existence of three different sequences of states which become clearly visible in the limit of the high density of states which can be reached by increasing the mass value parameter (Gilmore et al. 1986). Such qualitative modification of the energy spectrum of the 1D-quantum Hamiltonian gives the simplest example of the phenomenon which can be described as a quantum bifurcation. Figure 1 shows a schematic representation of quantum bifurcations for a model system with one degree-of-freedom in parallel in quantum and classical mechanics.

After looking for one simple example we can formulate a more general question which concerns the appearance in more general quantum systems of qualitative phenomena which can be characterized as quantum bifurcations.

## Simplest Effective Hamiltonians

We turn now to several models which describe some specific classes of relatively simple real

physical quantum systems formed by a finite number of particles (atoms, molecules, ...). Spectra of such quantum objects are studied nowadays with very high accuracy and this allows us to compare the behavior predicted by quantum bifurcations with the precise information about energy level structure found, for example, from high-resolution molecular spectroscopy.

Typically, the intra-molecular dynamics can be split into electronic, vibrational, and rotational ones due to important differences in characteristic energy excitations or in time scales. The most classical is the rotational motion and probably due to that the quantum bifurcations as a counterpart to classical bifurcations were first studied for purely rotational problems (Pavlichenkov 1993; Pavlichenkov and Zhilinskii 1988).

Effective rotational Hamiltonians describe the internal structure of rotational multiplets formed by isolated finite particle systems (atoms, molecules, nuclei) (Harter 1988). For many molecular systems in the ground electronic state any electronic and vibrational excitations are much more energy consuming as compared with rotational excitations. Thus, to study the molecular rotation the simplest physical assumption is to suppose that all electronic and all vibrational degrees-of-freedom are frozen. This means that a set of quantum numbers is given which have the sense of approximate integrals of motion specifying the character of vibrational and electronic motions in terms of these “good” quantum numbers. At the same time for a free molecule in the absence of any external fields due to isotropy of the space the total angular momentum  $J$  and its projection  $J_z$  on the laboratory fixed frame are strict integrals of motion. Consequently, to describe the rotational motion for fixed values of  $J$  and  $J_z$  it is sufficient to analyze the effective problem with only one degree-of-freedom. The dimension of classical phase space in this case equals two and the two classical conjugate variables are: the projection of the total angular momentum on the body fixed frame and conjugate angle variable. The classical phase space is topologically a two-dimensional sphere,  $S^2$ . There is a one-to-one correspondence between the points on a sphere and the orientation of the

angular momentum in the body-fixed frame. Such a representation gives a clear visualization of a classical rotational Hamiltonian as a function defined over a sphere (Harter 1988; Marsden and Ratiu 1994).

In quantum mechanics the rotation of molecules is traditionally described in terms of an effective rotational Hamiltonian which is constructed as a series in rotational operators  $J_x, J_y, J_z$ , the components of the total angular momentum  $\mathbf{J}$ . In a suitably chosen molecular fixed frame the effective Hamiltonian has the form

$$H_{\text{eff}} = AJ_x^2 + BJ_y^2 + CJ_z^2 + \sum c_{\alpha\beta\gamma} J_x^\alpha J_y^\beta J_z^\gamma + \dots, \quad (1)$$

where  $A, B, C$  and  $c_{\alpha\beta\gamma}$  are constants. To relate quantum and classical pictures we note that  $\mathbf{J}^2$  and energy are integrals of Euler's equations of motion for dynamic variables  $J_x, J_y, J_z$ . The phase space of the classical rotational problem with constant  $|\mathbf{J}|$  is  $S^2$ , the two-dimensional sphere, and it can be parametrized with spherical angles  $(\theta, \phi)$  in such a way that the points on  $S^2$  define the orientation of  $\mathbf{J}$ , i.e. the position of the axis and the direction of rotation. To get the classical interpretation of the quantum Hamiltonian we introduce the classical analogs of the operators  $J_x, J_y, J_z$

$$\mathbf{J} \rightarrow \begin{pmatrix} J_x \\ J_y \\ J_z \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix} \sqrt{J(J+1)} \quad (2)$$

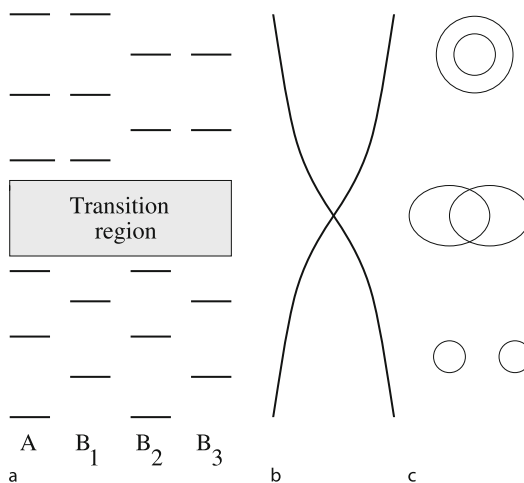
and consider the rotational energy as a function of the dynamical variables  $(\theta, \phi)$  and the parameter  $J$ .

Thus, for an effective rotational Hamiltonian the corresponding classical symbol is a function  $E_J(\theta, \phi)$  defined over  $S_2$  and named usually the rotational energy surface (Harter 1988).

Taking into account the symmetry imposed by the initial problem and the topology of the phase space the simplest rotational Hamiltonian can be constructed. In classical mechanics the simplest Hamiltonian can be defined (using Morse theory

(Morse 1925; Zhilinskiĭ 2001)) as a Hamiltonian function with the minimal possible number of non-degenerate stationary points compatible with the symmetry group action of the classical phase space. Morse theory in the presence of symmetry (or equivariant Morse theory) implies important restrictions on the number of minima, maxima, and saddle points. In the absence of symmetry the simplest Morse type function on the  $S^2$  phase space has one minimum and one maximum, as a consequence of Morse inequalities. In the presence of non-trivial symmetry group action the minimal number of stationary points on the sphere increases. For example, many asymmetric top molecules (possessing three different moment of inertia of the equilibrium configuration) have  $D_{2h}$  symmetry group (Landau and Lifschits 1981). This group includes rotations over  $\pi$  around  $\{x, y, z\}$  axes, reflections in  $\{xy, yz, zx\}$  planes and inversion asymmetry operations. Any  $D_{2h}$  invariant function on the sphere has at least six stationary points (two equivalent minima, two equivalent maxima, and two equivalent saddle points). This means that in quantum mechanics the asymmetric top has eigenvalues which form two regular sequences of quasi-degenerate doublets with the transition region between them. The correspondence between the quantum spectrum and the structure of the energy map for the classical problem is shown in Fig. 2. Highly symmetrical molecules which have cubic symmetry, for example, can be described by a simplest Morse-type Hamiltonian with 26 stationary points (6 and 8 minima/maxima and 12 saddle points). As a consequence, the corresponding quantum Hamiltonian shows the presence of six-fold and eight-fold quasi-degenerate clusters of rotational levels.

As soon as the simplest classical Hamiltonian is characterized by the appropriate system of stationary points the whole region of possible classical energy values (in the case of dynamical systems with only one degree-of-freedom the energy-momentum map becomes simply the energy map) appears to be split into different regions corresponding to different dynamical regimes, i.e. to different regions of the phase portrait foliated by topologically non-equivalent



**Quantum Bifurcations, Fig. 2** (a) Schematic representation of the energy level structure for asymmetric top molecule. Vertical axis corresponds to energy variation. Quantum levels are classified by the symmetry group of the asymmetric top. Two fold clusters at two ends of the rotational multiplet are formed by states with different symmetry. (b) Foliation of the classical phase space ( $S^2$  sphere) by constant levels of the Hamiltonian given in the form of its Reeb graph. Each point corresponds to a connected component of the constant level set of the Hamiltonian (energy). (c) Geometric representation of the constant energy sections

systems of classical trajectories. Accordingly, the energy spectrum of the corresponding quantum Hamiltonian can be qualitatively described as formed by regular sequences of states within each region of the classical energy map.

Quantum bifurcations are universal phenomena which lead to a new organization of the energy spectrum into qualitatively different regions in accordance with corresponding qualitative modifications of the classical energy-momentum map under the variation of some control parameter.

### Simplest Hamiltonians for Two Degree-of-Freedom Systems

When the quantum system has two or larger number of degrees-of-freedom the simplest dynamical regimes often correspond in classical mechanics to a quasi-regular dynamics which can be reasonably well approximated by an integrable model. The integrable model in classical mechanics can

be constructed by normalizing the Hamiltonian and by passing to so-called normal forms (Arnold 1989; Marsden and Ratiu 1994). The quantum counterpart of normalization is the construction of a mutually commuting set of operators which should not be mistaken with quantization of systems in normal form. Corresponding eigenvalues can be used as “good” quantum numbers to label quantum states. A joint spectrum of mutually commuting operators corresponds to the image of the energy-momentum map for the classical completely integrable dynamical problem. In this context the question about quantum bifurcations first of all leads to the question about qualitative classification of the joint spectra of mutually commuting operators. To answer this question we need to start with the qualitative description of foliations of the total phase space of the classical problem by common levels of integrals of motion which are mutually in involution (Arnold 1989; Bolsinov and Fomenko 2004). One needs to distinguish the regular and the singular values of the energy-momentum map. For Hamiltonian systems the inverse images of the regular values are regular tori (one or several) (Arnold 1989). A lot of different singularities are possible. In classical mechanics different levels of the classifications are studied in detail (Bolsinov and Fomenko 2004). The diagram which represents the image of the classical EM map together with its stratification into regular and critical values is named the bifurcation diagram. The origin of such a name is due to the fact that the values of integrals of motion can be considered as control parameters for the phase portraits (inverse images of the EM map) of the reduced systems.

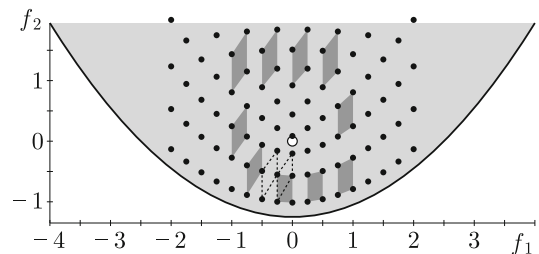
For quantum problems the analog of the classical stratification of the EM map for integrable systems is the splitting of the joint spectrum of several commuting observables into regions formed by regular lattices of joint eigenvalues. Any local simply connected neighborhood of a regular point of the lattice can be deformed into part of the regular  $\mathbb{Z}^n$  lattice of integers. This means that local quantum numbers can be consistently introduced to label states of the joint spectrum. If the regular region is not simply connected

it still can be characterized locally by a set of “good” quantum numbers. At the same time this is impossible globally. Likewise in classical mechanics the Hamiltonian monodromy is the simplest obstruction to the existence of the global action-angle variables (Duistermaat 1980; Nekhoroshev 1972), in quantum mechanics the analog notion of quantum monodromy (Cushman and Duistermaat 1988; Grondin et al. 2002; Sadovskii and Zhilinskii 1999; Vũ Ngoc 1999) characterizes the global non-triviality of the regular part of the lattice of joint eigenvalues. Figure 3 demonstrates the effect of the presence of a classical singularity (isolated focus-focus point) on the global properties of the quantum lattice formed by joint eigenvalues of two commuting operators for a simple problem with two degrees-of-freedom, which is essentially the  $1 : (-1)$  resonant oscillator (Nekhoroshev et al. 2006). Two integrals of motions in this example are chosen as

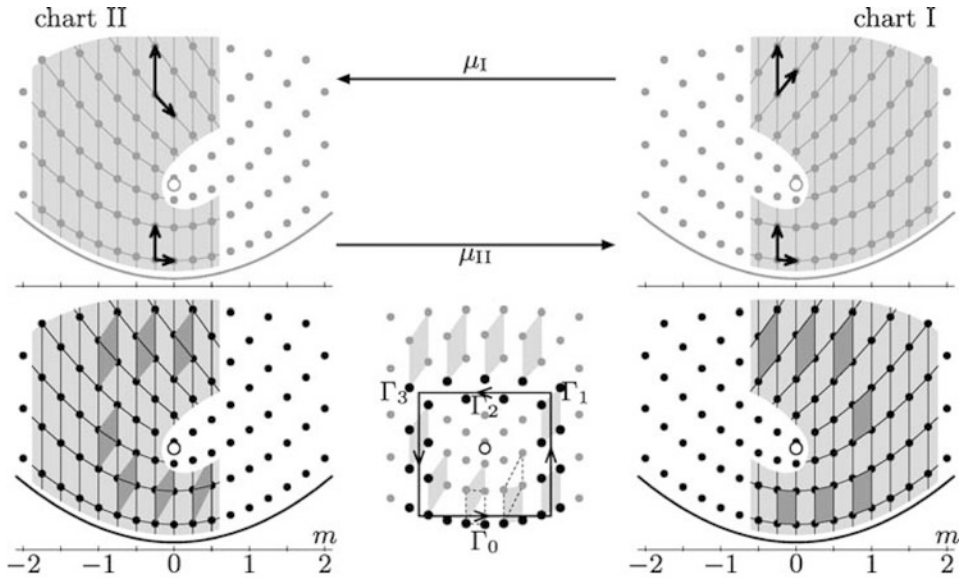
$$f_1 = \frac{1}{2} (p_1^2 + q_1^2) - \frac{1}{2} (p_2^2 + q_2^2), \quad (3)$$

$$f_2 = p_1 q_2 + p_2 q_1 + \frac{1}{4} (p_1^2 + q_1^2 + p_2^2 + q_2^2)^2. \quad (4)$$

Locally in any simply connected region which does not include the classical singularity of the EM map situated at  $f_1 = f_2 = 0$ , the joint spectrum



**Quantum Bifurcations, Fig. 3** Joint spectrum of two commuting operators together with the image of the classical EM map for the resonant  $1 : (-1)$  oscillator given by (3), (4). Quantum monodromy is seen as a result of transportation of the elementary cell of the quantum lattice along a close path through a non simply connected region of the regular part of the image of the EM map. (Taken from (Nekhoroshev et al. 2006))



**Quantum Bifurcations, Fig. 4** Two chart atlas which covers the quantum lattice of the  $1 : (-1)$  resonant oscillator system represented in Fig. 3. *Top plots* show the choice of basis cells and the gluing map between the charts. *Bottom plots* show the transport of the elementary cell

(dark gray quadrangles) in each chart. *Central bottom panel* shows closed path  $\Gamma$  and its quantum realization (black dots) leading to non-trivial monodromy (compare with Fig. 3). (Taken from (Nekhoroshev et al. 2006))

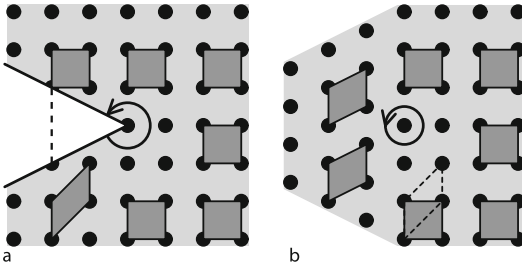
can be smoothly deformed to the regular  $Z^2$  lattice (Nekhoroshev et al. 2006; Zhilinskii 2006). Such lattices are shown, for example, in Fig. 4. If somebody wants to use only one chart to label states, it is necessary to take care in respect of the multi-valuedness of such a representation. There are two possibilities:

- (i) One makes a cut and maps the quantum lattice to a regular  $Z^2$  lattice with an appropriate solid angle removed from it (see Fig. 5 (Nekhoroshev et al. 2006; Sadovskii and Zhilinskii 1999; Zhilinskii 2006)). Points on the boundary of such a cut should be identified and a special matching rule explaining how to cross the path should be introduced. Similar constructions are quite popular in solid state physics in order to represent defects of lattices, like dislocations, disclinations, etc. We just note that the “monodromy defect” introduced in such a way is different from standard construction for dislocation and disclination defects (Kleman 1983; Mermin 1979). The inverse

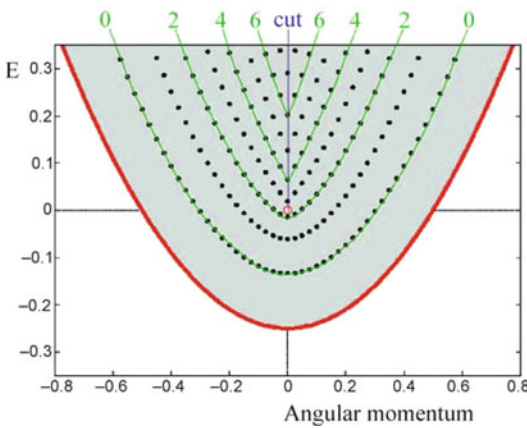
procedure of the construction of the “monodromy defect” (Zhilinskii 2006) from a regular lattice is represented in Fig. 5. Let us note that the width of the solid angle removed depends on the direction of the cut and the direction of the cut itself can be chosen in an ambiguous way.

- (ii) An alternative possibility is to make a cut in such a way that the width of the removed angle becomes equal to zero. For focus-focus singularities one such direction always exists and is named an eigenray by Symington (2003). The same construction is used in some physical papers (Child 2001; Child et al. 1999; Winnewisser et al. 2006). The inconvenience of such a procedure is the appearance of discontinuity of the slope of the constant action (quantum number) line at the cut, whereas the values of actions themselves are continued (see Fig. 6). This gives the wrong impression that this eigenray is associated with some special non-regular behavior of the initial problem, whereas there is no singularity except at one focus-focus point.





**Quantum Bifurcations, Fig. 5** Construction of the  $1 : (-1)$  lattice defect starting from the regular  $Z^2$  lattice. The solid angle is removed from the regular  $Z^2$  lattice and points on the so-obtained boundary are identified by vertical shifting. Dark gray quadrangles show the evolution of an elementary lattice cell along a closed path around the defect point. (Taken from (Nekhoroshev et al. 2006))



**Quantum Bifurcations, Fig. 6** Representation of the quantum joint spectrum for the “Mexican hat” potential  $V(r) = ar^4 - br^2$  with the “cut” along the eigenray. For such a cut the left and the right limits at the cut give the same values of actions (good quantum numbers) but the lines of constant values of actions exhibit a “kink” at the cut (the discontinuity of the first derivative)

## Bifurcations and Symmetry

The general mathematical answer about the possible qualitative modifications of a system of stationary points of functions depending on some control parameters can be found in bifurcation (or catastrophe) theory (Arnold 1992; Gilmore 1981; Golubitsky and Schaeffer 1984). It is important that the answer depends on the number of control parameters and on the symmetry. Very simple classification of possible typical bifurcations of stationary points of a one-parameter family of functions under presence of symmetry can

be formulated for dynamical systems with one degree-of-freedom. The situation is particularly simple because the phase space is two-dimensional and the complete list of local symmetry groups (which are the stabilizers of stationary points) includes only 2D-point groups (Weyl 1952). It should be noted that the global symmetry of the problem can be larger than the local symmetry of the bifurcating stationary points. In such a case the bifurcations occur simultaneously for all points forming one orbit of the global symmetry group (Michel 1980; Michel and Zhilinskii 2001a). We describe briefly here (see Table 1) the classification of the bifurcations of stationary points in the presence of symmetry for families of functions depending on one parameter and associated quantum bifurcations (Pavlichenkov 1993; Pavlichenkov and Zhilinskii 1988). Their notation includes the local symmetry group and several additional indexes which specify creation/annihilation of stationary points and the local or non-local character of the bifurcation. The list of possible bifurcations includes:

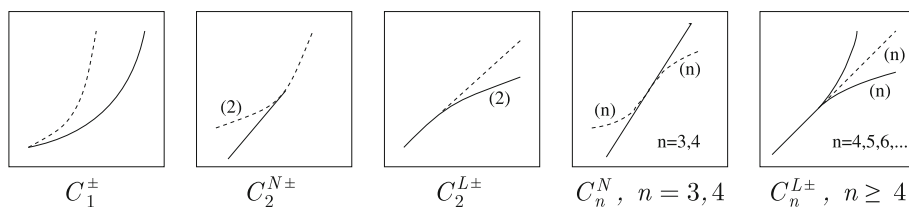
$C_1^\pm$  A non-symmetrical non-local bifurcation resulting in the appearance (+) or disappearance (−) of a stable-unstable pair of stationary points with the trivial local symmetry  $C_1$ . In the quantum problem this bifurcation is associated with the appearance or disappearance of a new regular sequence of states glued at its end with the intermediate part of another regular sequence of quantum states (Gilmore et al. 1986; Uwano 1999).

$C_2^{L\pm}$  A local bifurcation with the broken  $C_2$  local symmetry. This bifurcation results either in appearance of a triple of points (two equivalent stable points with  $C_1$  local symmetry and one unstable point with  $C_2$  local symmetry) instead of one stable point with  $C_2$  symmetry, or in inverse transformation. The number of stationary points in this bifurcation increases or decreases by two. For the quantum problem the result is the transformation of a local part of a regular sequence of states into one sequence of quasi-degenerate doublets.

$C_2^{N\pm}$  A non-local bifurcation with the broken  $C_2$  local symmetry. This bifurcation results in



**Quantum Bifurcations, Table 1** Bifurcations in the presence of symmetry. *Solid lines* denote stable stationary points. *Dashed lines* denote unstable stationary points. *Numbers in parenthesis* indicate the multiplicity of stationary points



appearance (+) or disappearance (−) of two new unstable points with broken  $C_2$  symmetry and simultaneous transformation of the initially stable (for +) or unstable (for −) stationary point into an unstable/stable one. The number of stationary points in this bifurcation increases or decreases by two. For the quantum problem this means the appearance of a new regular sequence of states near the separatrix between two different regular regions.

$C_n^N$  ( $n = 3, 4$ ) A non-local bifurcation corresponding to passage of  $n$  unstable stationary points through a stable stationary point with  $C_n$  local symmetry which is accompanied with the minimum  $\leftrightarrow$  maximum change for a stable point with the  $C_n$  local symmetry. The number of stationary points remains the same. For the quantum problem this bifurcation corresponds to transformation of the increased sequence of energy levels into a decreased sequence.

$C_n^{L\pm}$  ( $n \geq 4$ ) A local bifurcation which results in appearance (+) or disappearance (−) of  $n$  stable and  $n$  unstable stationary points with the broken  $C_n$  symmetry and a simultaneous minimum  $\leftrightarrow$  maximum change of a stable point with the  $C_n$  local symmetry. The number of stationary points increases or decreases by  $2n$ . In the quantum problem after bifurcation a new sequence of  $n$ -times quasi-degenerate levels appears/disappears.

Universal quantum Hamiltonians which describe the qualitative modification of the quantum energy level system around the bifurcation point are given in (Pavlichenkov 1993; Pavlichenkov and Zhilinskiĭ 1988).

The presence of symmetry makes it much easier to observe the manifestation of quantum

bifurcations. Modification of the local symmetry of stable stationary points results in the modification of the cluster structure of energy levels, i.e. the number and the symmetry types of energy level forming quasi-degenerate groups of levels. This phenomenon is essentially the spontaneous breaking of symmetry (Michel 1980). Several concrete molecular systems which show the presence of quantum bifurcations in rotational structure under rotational excitation are cited in Table 2. Many other examples can be found in (Child 2000; Efsthathiou et al. 2004c; Pavlichenkov 1993; Sadovskii and Zhilinskiĭ 1993; Sadovskii et al. 1990; Zhilinskiĭ 1996, 2001; Zhilinskiĭ and Pavlichenkov 1988; Zhilinskiĭ and Petrov 1996) and references therein. In purely vibrational problems breaking dynamical  $SU(N)$  symmetry of the isotrope harmonic oscillator till finite symmetry group results in formation of so-called non-linear normal modes (Efsthathiou et al. 2004c; Montaldi et al. 1988) or quasimodes (Arnold 1972), or local modes (Child 2000; Ezra 1996; Joyeux et al. 2002, 2005; Kellman 1995; Kozin et al. 2005; Lu and Kellman 1997). In the case of two degrees-of-freedom the analysis of the vibrational problem can be reduced to the analysis of the problem similar to the rotational one (Harter 1988; Sadovskii et al. 1993) and all the results about possible types of bifurcations found for rotational problems remain valid in the case of intra-molecular vibrational dynamics.

### Imperfect Bifurcations

According to general results the possible types of bifurcations which are generically present (and persist under small deformations) in a family of dynamical systems strictly depend on the number of control parameters. In the absence of symmetry

**Quantum Bifurcations, Table 2** Molecular examples of quantum bifurcations in the rotational structure of individual vibrational components under the variation of the absolute value of angular momentum,  $J_c$  is the critical value corresponding to bifurcation

| Molecule          | Component   | $J_c$ | Bifurcation type                   |
|-------------------|-------------|-------|------------------------------------|
| SiH <sub>4</sub>  | $v_2(+)$    | 12    | $C_2^{N+}$                         |
| SnH <sub>4</sub>  | $v_2(-)$    | 10    | $C_2^{N+}, C_3^N, C_4^N, C_2^{N-}$ |
| CF <sub>4</sub>   | $v_2(+)$    | 50    | $C_4^{L+}$                         |
| H <sub>2</sub> Se | $ 0\rangle$ | 20    | $C_2^{L+}$                         |

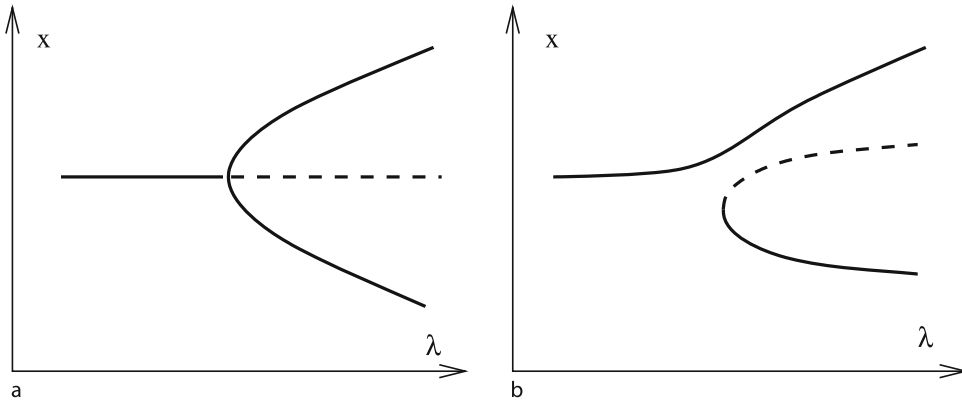
only one bifurcation of stationary points is present for a one-parameter family of Morse-type functions, namely the formation (annihilation) of two new stationary points. This corresponds to saddle-node bifurcation for one degree-of-freedom Hamiltonian systems. The presence of symmetry increases significantly the number of possible bifurcations even for families with only one parameter (Gilmore 1981; Golubitsky and Schaeffer 1984). From the physical point-of-view it is quite natural to study the effect of symmetry breaking on the symmetry allowed bifurcation. Decreasing symmetry naturally results in the modification of the allowed types of bifurcations but at the same time it is quite clear that at sufficient slight symmetry breaking perturbation the resulting behavior of the system should be rather close to the behavior of the original system with higher symmetry.

In the case of a small violation of symmetry the so-called “imperfect bifurcations” can be observed. Imperfect bifurcations, which are well known in the classical theory of bifurcations (Golubitsky and Schaeffer 1984) consist of the appearance of stationary points in the neighborhood of another stationary point which does not change its stability. In some way one can say that imperfect bifurcation mimics generic bifurcation in the presence of higher symmetry by the special organization of several bifurcations which are generic in the presence of lower symmetry. Naturally quantum bifurcations follow the same behavior under the symmetry breaking as classical ones. Very simple and quite natural examples of imperfect quantum bifurcations were demonstrated on the example of the rotational structure modifications under increasing angular momentum (Zhilinskii et al.

1999). The idea of appearance of imperfect bifurcations is as follows. Let us suppose that some symmetrical molecule demonstrates under the variation of angular momentum a quantum rotational bifurcation allowed by symmetry. The origin of this bifurcation is due, say, to centrifugal distortion effects which depend strongly on  $J$  but are not very sensitive to small variation of masses even in the case of symmetry breaking isotopic substitution. In such a case a slight modification of the masses of one or several equivalent atoms breaks the symmetry and this symmetry violation can be made very weak due to the small ratio  $\Delta M/M$  under isotope substitution. In classical theory the effect of symmetry breaking can be easily seen through the variation of the position of stationary points with control parameter. For example, instead of a pitchfork bifurcation which is typical for  $C_2$  local symmetry, we get for the unsymmetrical problem (after slight breaking of  $C_2$  symmetry) a smooth evolution of the position of one stationary point and the appearance of two new stationary points in fold catastrophe (see Fig. 7). In associated quantum bifurcations the most important effect is the splitting of clusters. But one should be careful with this interpretation because in quantum mechanics of finite particle systems the clusters are always split due to quantum mechanical tunneling between different equivalent regions of localization of quantum wave functions. Intercluster splitting increases rapidly approaching the region of classical separatrix. The behavior of quantum tunneling was studied extensively in relation to the quantum breathers problem (Aubry et al. 1996; Flach and Willis 1998). Systematic application of quasi-classical methods to reproduce quantum energy level structure near the singularities of the energy-momentum maps where exponentially small corrections are important is possible but requires special efforts (see for example (Colin de Verdier and Vũ Ngoc 2003)) and we will not touch upon this problem here.

### Organization of Bifurcations

The analysis of the quantum bifurcations in concrete examples of rotating molecules have shown that in some cases the molecule undergoes several consecutive qualitative changes which can be



**Quantum Bifurcations, Fig. 7** Imperfect bifurcations. (a) Position  $x$  of stationary points as a function of control parameter  $\lambda$  during a pitchfork bifurcation in the presence of  $C_2$  local symmetry. (b) Modifications induced by small

symmetry perturbation of lower symmetry. *Solid line*: Stable stationary points. *Dashed lines*: Unstable stationary points

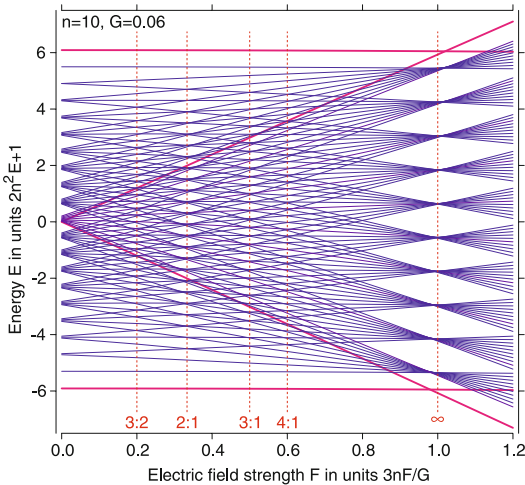
interpreted as a sequence of bifurcations which sometimes cannot even be separated into elementary bifurcations for the real scale of the control parameter (Zhilinskiĭ 2001). One can imagine in principle that successive bifurcations lead to quantum chaos in analogy with classical dynamical systems where the typical scenario for the transition to chaos is through a sequence of bifurcations. Otherwise, the molecular examples were described with effective Hamiltonians depending only on one degree-of-freedom and the result of the sequence of bifurcations was just the crossover of the rotational multiplets (Pierre et al. 1989). In some sense such a sequence of bifurcations can be interpreted as an imperfect bifurcation assuming initially higher dynamical symmetry, like the continuous  $SO(3)$  group.

Later, a similar crossover phenomenon was found in a quite different quantum problem, like the hydrogen atom in external fields (Efsthathiou et al. 2007b; Michel and Zhilinskiĭ 2001b; Sadovskii et al. 1996). The general idea of such organization of bifurcations is based on the existence of two different limiting cases of dynamical regimes for the same physical quantum system (often under presence of the same symmetry group) which are qualitatively different. For example, the number of stationary points, or their stability differs. If  $H_1$  and  $H_2$  are two corresponding effective Hamiltonians, the natural question is: Is it possible to transform  $H_1$  into  $H_2$

by a generic perturbation depending on only one parameter? And if so, what is the minimal number of bifurcations to go through?

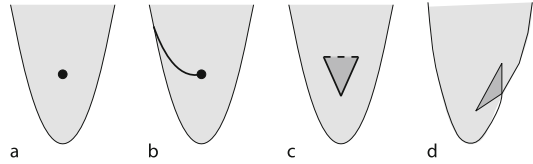
The simplest quantum system for which such a question becomes extremely natural is the hydrogen atom in the presence of external static electric ( $F$ ) and magnetic ( $G$ ) fields. Two natural limits – the Stark effect in the electric field and Zeeman effect in the magnetic field – show quite different qualitative structure even in the extremely low field limit (Cushman and Sadovskii 2000; Efsthathiou et al. 2004a; Peters et al. 1997; Sadovskii et al. 1996; Uzer 1990). Keeping a small field one can go from one (Stark) limit to another (Zeeman) and this transformation naturally goes through qualitatively different regimes (Efsthathiou et al. 2007b; Michel and Zhilinskiĭ 2001b). In spite of the fact that the hydrogen atom (even without spin and relativistic corrections) is only a three degree-of-freedom system, the complete description of qualitatively different regimes in a small field limit is still not done and remains an open problem (Efsthathiou et al. 2007b).

An example of clearly seen qualitative modifications of the quantum energy level system of the hydrogen atom under the variation of  $F/G$  ratio of the strengths of two parallel electric and magnetic fields is shown in Fig. 8. The calculations are done for a two degree-of-freedom system after the normalization with respect to the global action. In



**Quantum Bifurcations, Fig. 8** Reorganization of the internal structure of the  $n$ -multiplet of the hydrogen atom in small parallel electric and magnetic fields. Energies of stationary points of the classical Hamiltonian (red solid lines) are shown together with quantum energy levels (blue solid lines). The figure is done for  $n = 10$  (there are  $n^2 = 100$  energy levels forming this multiplet). As the ratio  $F/G$  of electric  $F$  and magnetic  $G$  fields varies this two degree-of-freedom system goes through different zones associated with special resonance relations between two characteristic frequencies (shown by vertical dashed lines). (Taken from (Efsthathiou et al. 2007b))

quantum mechanics language this means that only energy levels which belong to the same  $n$ -shell of the hydrogen atom are treated and the interaction with other  $n'$  shells is taken into account only effectively. The limiting classical phase space for this effective problem is the four-dimensional space  $S^2 \times S^2$ , which is the direct product of two two-dimensional spheres. In the presence of axial symmetry this problem is completely integrable and the Hamiltonian and the angular momentum provide a complete set of mutually commuting operators. Energies of stationary points of classical Hamiltonian limit are shown on the same Fig. 8 along with quantum levels. When one of the characteristic frequencies goes through zero, the so-called collapse phenomena occurs. Some other non-trivial resonance relations between two frequencies are also indicated. These resonances correspond to special organization of quantum energy levels. At the same time it is not necessary here to go to joint spectrum representation in order to see the reorganization of stationary points of the



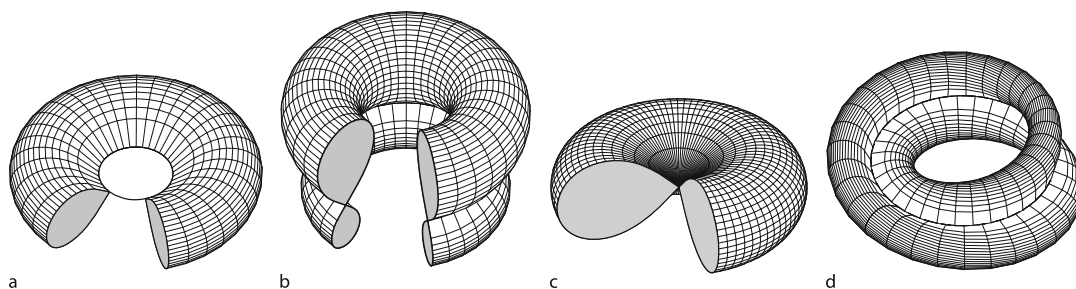
**Quantum Bifurcations, Fig. 9** Typical images of the energy momentum map for completely integrable Hamiltonian systems with two degrees-of-freedom in the case of: (a) integer monodromy, (b) fractional monodromy, (c) non-local monodromy, and (d) bidromy. Values in the light shaded area lift to single 2-tori; values in the dark shaded area lift to two 2-tori. (Taken from (Sadovskii and Zhilinskii 2006))

Hamiltonian function on  $S^2 \times S^2$  phase space under the variation of the external control parameter  $F/G$ . A more detailed treatment of qualitative features of the energy level systems for the hydrogen atom in low fields is given in (Cushman and Sadovskii 2000; Efsthathiou et al. 2004a, 2007b).

## Bifurcation Diagrams for Two Degree-of-Freedom Integrable Systems

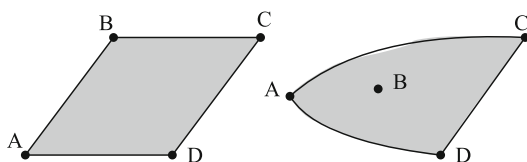
Let us consider now the two degree-of-freedom integrable system with compact phase space as a bit more complex but still reasonably simple problem. Many examples of such systems possess EM maps with the stratification of the image formed by the regular part surrounded by the singular boundary. The most naturally arising examples of classical phase spaces, like  $S^2 \times S^2$ ,  $CP^2$ , are of that type. All internal points on the image of the EM map are regular in these cases. In practice, real physical problems, even after necessary simplifications and approximations lead to more complicated models. Some examples of fragments of images of the EM map with internal singular points are shown in Fig. 9. In classical mechanics the inverse images of critical values are singular tori of different kinds. Some of them are represented in Fig. 10. Inverse images of critical points situated on the boundary of the EM image have lower dimension. They can be one-dimensional tori ( $S^1$ -circles), or zero-dimensional (points).

The natural question now is to describe typical generic modifications of the Hamiltonian which



**Quantum Bifurcations, Fig. 10** Two-dimensional singular fibers in the case of integrable Hamiltonian systems with two degrees-of-freedom (*left to right*): singular torus, bitorus, pinched and curled tori. Singular torus corresponds to critical values in Fig. 9c, d (ends of bitoris line). Bitorus corresponds to critical values in Fig. 9c, d, which belong to

singular line (fusion of two components). Pinched torus corresponds to isolated focus-focus singularity in Fig. 9a. Curled torus is associated with critical values at singular line in Fig. 9b (fractional monodromy). (Taken from (Sadovskii and Zhilinskii 2006))



**Quantum Bifurcations, Fig. 11** Qualitative modification of the image of the EM map due to Hamiltonian Hopf bifurcation. *Left*: Simplest integrable toric fibration over  $S^2 \times S^2$  classical phase space. *A, B, C, D*: Critical values corresponding to singular  $S^0$  fibers. Regular points on the boundary correspond to  $S^1$  fibers. Regular internal points: Regular  $T^2$  fibers. *Right*: Appearance of an isolated critical value inside the field of regular values. Critical value *B* corresponds to pinched torus shown in Fig. 10

lead to qualitative modifications of the EM map image in classical mechanics and to associated modifications of the joint spectrum in quantum mechanics.

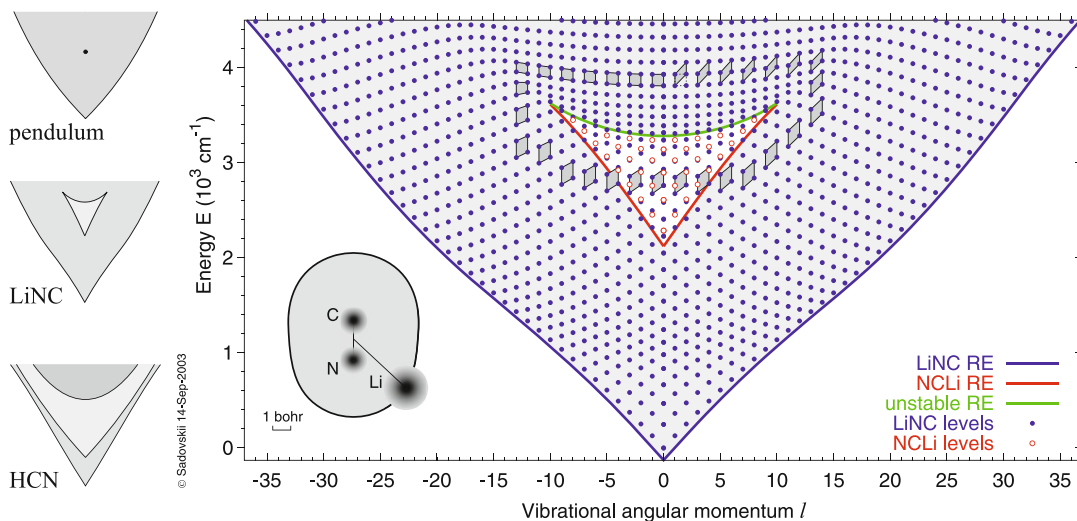
The simplest classical bifurcation leading to modification of the image of the EM map is the Hamiltonian Hopf bifurcation (Van der Meer 1985). It is associated with the following modification of the image of the EM map. The critical value of the EM map situated on the boundary leaves the boundary and enters the internal domain of regular values (see Fig. 11). As a consequence, the toric fibration over the closed path surrounding an isolated singularity is non-trivial. Its non-triviality can be characterized by the Hamiltonian monodromy which describes the mapping from the fundamental group of the base space into the first homology group of the regular fiber (Duistermaat 1998). A typical pattern of the joint

spectrum around such a classical singularity is shown in Fig. 3. The joint spectrum manifests the presence of quantum monodromy. Its interpretation in terms of regular lattices is given in Figs. 4 and 5.

Taking into account additional terms of higher order it is possible to distinguish different types of Hamiltonian Hopf bifurcations usually named as subcritical and supercritical (Efsthathiou 2004; Van der Meer 1985). New qualitative modification, for example, corresponds to transformation of an isolated singular value of the EM map into an “island”, i.e. the region of the EM image filled by points whose inverse images consist of two connected components. Integrable approximation for vibrational motion in the LiCN molecule shows the presence of such an island associated with the non-local quantum monodromy (see Fig. 12) (Joyeux et al. 2003). The monodromy naturally coincides with the quantum monodromy of isolated focus-focus singularity which deforms continuously into the island monodromy. It is interesting to note that in molecule HCN which is rather similar to LiCN, the region with two components in the inverse image of the EM map exists also but the monodromy cannot be defined due to impossibility to go around the island (Efsthathiou et al. 2004b).

In the quantum problem the presence of “standard” quantum monodromy in the joint spectrum of two mutually commuting observables can be seen through the mapping of a locally regular part of the joint spectrum lattice to an idealized  $Z^2$





**Quantum Bifurcations, Fig. 12** Quantum joint spectrum for the quantum model problem with two degrees-of-freedom describing two vibrations in the LiCN molecule. The non-local quantum monodromy is shown by the evolution of the elementary cell of the quantum lattice around the singular line associated with gluing of two regular lattices corresponding in molecular language to

two different isomers, LiCN and LiNC. Classical limit (*left*) shows the possible deformation of isolated focus-focus singularity for pendulum to non-local island singularity for LiNC model. In contrast to LiCN, the HCN model has an infinite island which cannot be surrounded by a close path. (Taken from (Joyeux et al. 2003))

lattice. Existence of local actions for the classical problem which are defined almost everywhere and the multivaluedness of global actions from one side and the quantum-classical correspondence from another side allow the interpretation of the joint spectrum with quantum monodromy as a regular lattice with an isolated defect.

Recently, the generalization of the notion of quantum (and classical) monodromy was suggested (Efsthathiou et al. 2007a; Nekhoroshev et al. 2006). For quantum problems the idea is based on the possibility to study instead of the complete lattice formed by the joint spectrum only a sub-lattice of finite index. Such a transformation allows one to eliminate certain “weak line singularities” presented in the image of the EM map. The resulting monodromy is named “fractional monodromy” because for the elementary cell in the regular region the formal transformation after a propagation along a close path crossing “weak line singularities” turns out to be represented in a form of a matrix with fractional coefficients.

An example of quantum fractional monodromy can be given with a  $1 : (-2)$  resonant oscillator system possessing two integrals of motion  $f_1, f_2$  in involution:

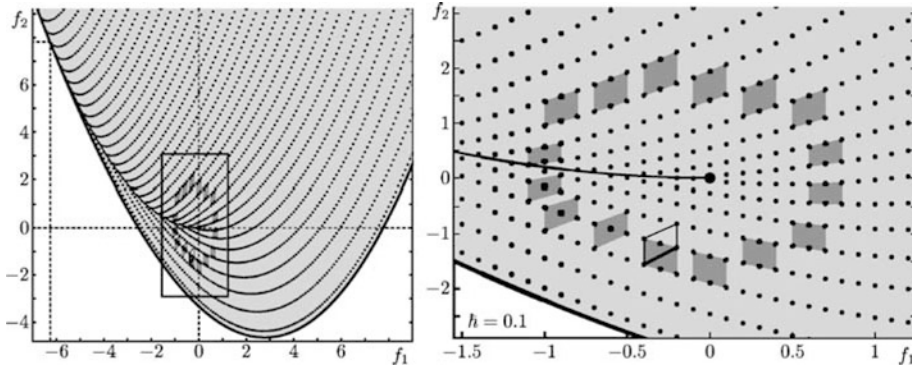
$$f_1 = \frac{\omega}{2} (p_1^2 + q_1^2) - \frac{2\omega}{2} (p_2^2 + q_2^2) + R_1(q, p), \quad (5)$$

$$f_2 = \text{Im} \left[ (q_1 + ip_1)^2 (q_2 + ip_2) \right] + R_2(q, p). \quad (6)$$

The corresponding joint spectrum for the quantum problem is shown in Fig. 13. It can be represented as a regular  $Z^2$  lattice with a solid angle removed (see Fig. 14). The main difference with the standard integer monodromy representation is due to the fact that even after gluing two sides of the cut we get the one-dimensional singular stratum which can be neglected only after going to a sub-lattice (to a sub-lattice of index 2 for  $1 : 2$  fractional singularity).

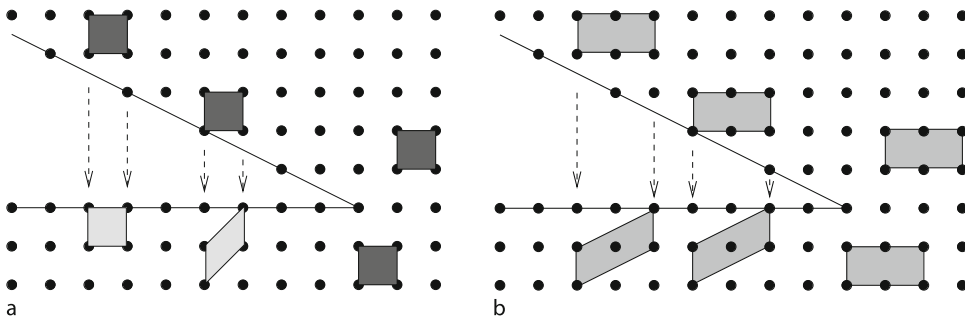
Another kind of generalization of the monodromy notion is related to the appearance of multi-component inverse images for the EM maps. We have already mentioned such a possibility with the appearance of non-local monodromy and Hamiltonian Hopf bifurcations (see Fig. 12). But in this case two components of the inverse image belong to different regular





**Quantum Bifurcations, Fig. 13** Joint quantum spectrum for two-dimensional non-linear 1 : (-2) resonant oscillator (5), (6). The singular line is formed by critical values whose inverse images are curled tori shown in

Fig. 10. In order to get the unambiguous result of the propagation of the cell of the quantum lattice along a closed path crossing the singular line, the elementary cell is doubled. (Taken from (Nekhoroshev et al. 2006))

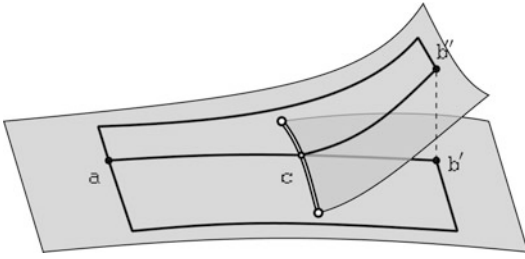


**Quantum Bifurcations, Fig. 14** Representation of a lattice with 1 : 2 rational defect by cutting and gluing. *Left*: The elementary cell goes through cut in an ambiguous

way. The result depends on the place where the cell crosses the cut. *Right*: Double cell crosses the cut in an unambiguous way. (Taken from (Nekhoroshev et al. 2006))

domains and cannot be joined by a path going only through regular values. Another possibility is suggested in (Sadovskii and Zhilinskii 2006, 2007) and is explained schematically in Fig. 15. This figure shows that the arrangement of fibers can be done in such a way that one connected component can be deformed into another connected component along a path which goes only through regular tori. The existence of a quantum joint spectrum corresponding to such a classical picture was demonstrated on the example of a very well-known model problem with three degrees-of-freedom: Three resonant oscillators with 1 : 1 : 2 resonance, axial symmetry and with small detuning between double degenerate and non-degenerate modes (Giacobbe et al. 2004; Sadovskii and Zhilinskii 2007). The specific behavior of the joint spectrum for this model can be characterized as

self-overlapping of a regular lattice. The possibility to propagate the initially chosen cell through a regular lattice from the region of self-overlapping of lattice back to the same region but to another component was named “bidromy”. More complicated construction for the same problem allows us to introduce the “bipath” notion. The bipath starts at a regular point of the EM image, and crosses the singular line by splitting itself into two components. Each component belongs to its proper lattice in the self-overlapping region. Two components of the path can go back through the regular region only and fuse together. The behavior of quantum cells along a bipath is shown in Fig. 16. Providing a rigorous mathematical description of such a construction is still an open problem. Although the original problem has three degrees-of-freedom, it is possible to construct



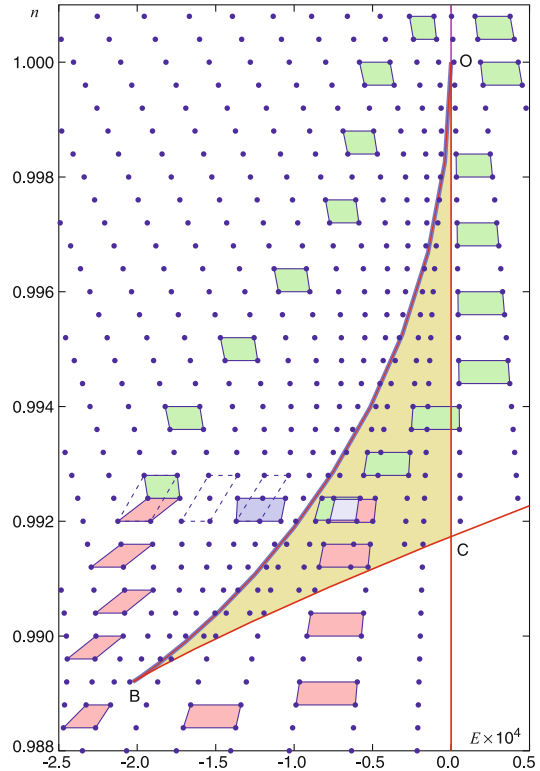
**Quantum Bifurcations, Fig. 15** Schematic representation of the inverse images for a problem with bidromy in the form of the unfolded surface. Each connected component of the inverse image is represented as a *single point*. The path  $b' - a - b''$  starts and ends at the same point of the space of possible values of integrals of motion but it starts at one connected component and ends at another one. At the same time the path goes only through regular tori. (Taken from (Sadovskii and Zhilinskii 2007))

a model system with two degrees-of-freedom and with similar properties.

### Bifurcations of “Quantum Bifurcation Diagrams”

We want now to stress some differences in the role of internal and external control parameters. From one point-of-view a quantum problem, which corresponds in the classical limit to a multidimensional integrable classical model, possesses a joint spectrum qualitatively described by a “quantum bifurcation diagram”. This diagram shows that the joint spectrum is formed from several parts of regular lattices through a cutting and gluing procedure. Going from one regular region to another is possible by crossing singular lines. The parameter defined along such a path can be treated as an internal control parameter. It is essentially a function of values of integrals of motion. To cross the singular line is equivalent to passing the quantum bifurcation for a family of reduced systems with a smaller number of degrees of freedom.

On the other side we can ask the following more general question. What kinds of generic modifications of “bifurcation diagrams” are possible for a family of integrable systems depending on some external parameters? Hamiltonian Hopf bifurcation leading to the appearance of a new isolated singular value and as a consequence



**Quantum Bifurcations, Fig. 16** Joint quantum spectrum for problem with bidromy. Quantum states are given by two numbers (energy,  $E$ , and polyad number,  $n$ ) which are the eigenvalues of two mutually commuting operators. Inside the  $OAB$  curvilinear triangle two regular lattices are clearly seen. One can be continued smoothly through the  $OC$  boundary whereas another continues through the  $BC$  boundary. This means that the regular part of the whole lattice can be considered as a one self-overlapping regular lattice. The figure suggests also the possibility to define the propagation of a double cell along a “bipath” through the singular line  $BO$  which leads to splitting of the cell into two elementary cells fusing at the end into one cell defining in such a way the “bidromy” transformation associated with a bipath. (Taken from (Sadovskii and Zhilinskii 2007))

appearance of monodromy is just one of the possible effects of this kind. Another possibility is the transformation of an isolated focus-focus singular value into the island associated with the presence of a second connected component of the inverse image of the EM map. It is also possible that such an island is born within the regular region of the EM map. In such a case naturally the monodromy transformation associated with a closed path surrounding the so-obtained island should be trivial (identity).

The boundary of the image of the EM map can also undergo transformation which results in the appearance of the region with two components in the inverse image but, in contrast to the previous example of the appearance of an island, these two components can be smoothly deformed one onto another along a continuous path going only through regular values of the EM map. Examples of all such modifications were studied on simple models inspired by concrete quantum molecular systems like the H atom, CO<sub>2</sub>, LiCN molecules and soon (Efsthathiou et al. 2007b; Giacobbe et al. 2004; Joyeux et al. 2003).

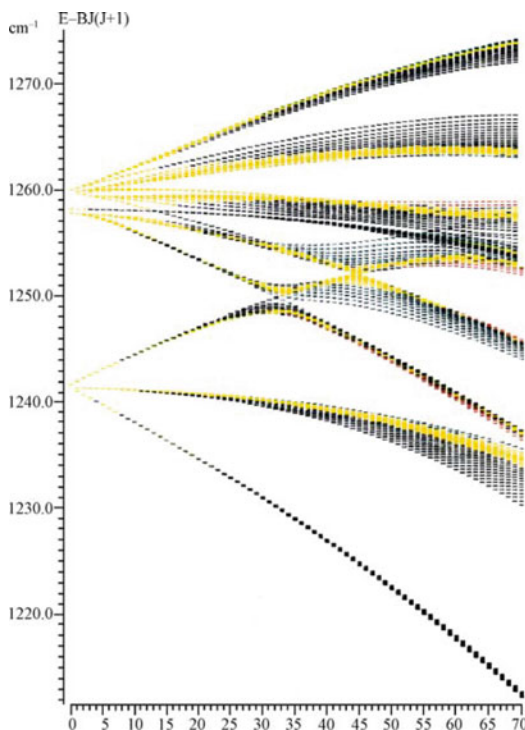
### Semi-Quantum Limit and Reorganization of Quantum Bands

Up to now we have discussed the qualitative modifications of internal structures of certain groups of quantum levels which are typically named bands. Their appearance is physically quite clear in the adiabatic approximation. The existence of fast and slow classical motions manifests itself in quantum mechanics through the formation of so-called energy bands. The big energy difference between energies of different bands correspond to fast classical variables whereas small energy differences between energy levels belonging to the same band correspond to classical slow variables. Typical bands in simple quantum systems correspond to vibrational structure of different electronic states, rotational structure of different vibrational states, etc.

If now we have a quantum problem which shows the presence of bands in its energy spectrum, the natural generalization consists of putting this quantum system in a family, depending on one (or several) control parameters. What are the generic qualitative modifications which can be observed within such a family of systems when control parameters vary? Apart from qualitative modifications of the internal structure of individual bands which can be treated as the earlier discussed quantum bifurcations, another qualitative phenomenon is possible, namely the redistribution of energy levels between bands or more generally, the reorganization of bands under the

variation of some control parameters (Brodersen and Zhilinskiĭ 1995; Faure and Zhilinskiĭ 2000, 2002; Pavlov-Verevkin et al. 1988; Sadovskii and Zhilinskiĭ 1999). In fact this phenomenon is very often observed in both the numerical simulations and the real experiments with molecular systems exhibiting bands. A typical example of molecular rovibrational energy levels classified according to their energy and angular momentum is shown in Fig. 17. It is important to note that the number of energy levels in bands before and after their “intersection” changes.

The same phenomenon of the redistribution of energy levels between energy bands can be understood by the example of a much simpler quantum system of two coupled angular momenta, say



**Quantum Bifurcations, Fig. 17** System of rovibrational energy levels of <sup>13</sup>CF<sub>4</sub> molecule represented schematically in  $E, J$  coordinates. The number of energy levels in each clearly seen band is  $2J + 1 + \delta$ , where  $\delta$  is a small integer which remains constant for isolated bands and changes at band intersections. In the semi-quantum model  $\delta$  is interpreted as the first Chern class, characterizing the non-triviality of the vector bundle formed by eigenfunctions of the “fast” subsystem over the classical phase space of the “slow” subsystem (Faure and Zhilinskiĭ 2001)

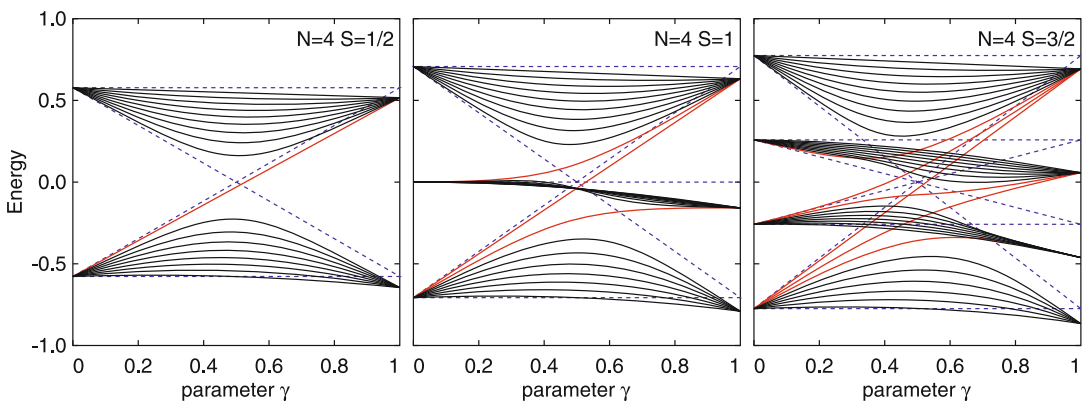
orbital angular momentum and spin in the presence of a magnetic field interacting only with spin (Pavlov-Verevkin et al. 1988; Sadovskii and Zhilinskiĭ 1999).

$$H = \frac{1-\gamma}{S} S_z + \frac{\gamma}{NS} (\mathbf{N} \cdot \mathbf{S}), \quad 0 \leq \gamma \leq 1. \quad (7)$$

The Hamiltonian for such a system can be represented in the form of a one-parameter family (7) having two natural limits corresponding to uncoupled and coupled angular momenta. The interpolation of eigenvalues between these two limits is shown in Fig. 18 for different values of spin quantum number,  $S = 1/2, 1, 3/2$ . The quantum number of orbital momentum is taken to be  $N = 4$ . Although this value is not much larger than the  $S$  values, the existence of bands and their reorganization under the variation of the external parameter  $\gamma$  is clearly seen in the figure.

Although the detailed description of this reorganization of bands will take us rather far away from the principal subject it is important to note that in the simplest situations there exists a very close relation between the redistribution phenomenon and the Hamiltonian Hopf bifurcations leading to the appearance of Hamiltonian monodromy (Vũ Ngoc 2007). In the semi-quantum limit when part of the dynamical variables are treated as purely classical and all the rest as quantum, the description of the complete system naturally leads

to a fiber bundle construction (Faure and Zhilinskiĭ 2001). The role of the base space is taken by the classical phase space for classical variables. A set of quantum wave-functions associated with one point of the base space forms a complex fiber. As a whole the so-obtained vector bundle with complex fibers can be topologically characterized by its rank and Chern classes (Nakahara 1990). Chern classes are related to the number of quantum states in bands formed due to quantum character of the total problem with respect to “classical” variables. Modification of the number of states in bands can occur only at band contact and is associated with the modification of Chern classes of the corresponding fiber bundle (Faure and Zhilinskiĭ 2000). The simplest situation takes place when the number of degrees of freedom associated with classical variables is one. In this case only one topological invariant – the first Chern class is sufficient to characterize the non-triviality of the fiber bundle and the difference in Chern classes is equal to the number of energy levels redistributed between corresponding bands. Moreover, in the generic situation (in the absence of symmetry) the typical behavior consists of the redistribution of only one energy level between two bands. The generic phenomena become more complicated with increasing the number of degrees of freedom for the classical part of variables. The model problem with two slow degrees of freedom (described in classical

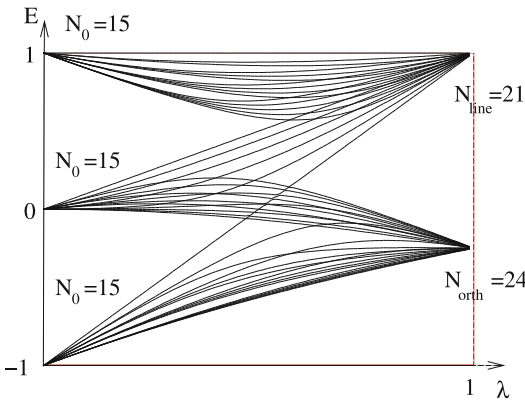


**Quantum Bifurcations, Fig. 18** Rearrangement of energy levels between bands for model Hamiltonian (7) with two, three, or four states for “fast” variable. Quantum

energy levels are shown by *solid lines*. Classical energies of stationary points for energy surfaces are shown by *dashed lines*. (Taken from (Sadovskii and Zhilinskiĭ 1999))

limit by the  $CP^2$  phase space) and three quantum states was studied in (Faure and Zhilinskii 2002). A new qualitative phenomenon was found, namely, the modification of the number of bands due to formation of topologically coupled bands. Figure 19 shows the evolution of the system of energy levels along with the variation of control parameter  $\lambda$ . Three quantum bands (at  $\lambda = 0$ ) transform into two bands (in the  $\lambda = 1$ ) limit. One of these bands has rank one, i.e. it is associated with one quantum state. Another has rank two. It is associated with two quantum states. Both bands have non-trivial topology (non-trivial Chern classes). Moreover, it is quite important that the newly formed topologically coupled band of rank two can be split into two bands of rank one only if a coupling with the third band is introduced.

The corresponding qualitative modifications of quantum spectra can be considered as natural generalizations of quantum bifurcations and probably should be treated as topological bifurcations. Thus, the description of possible “elementary” rearrangements of energy bands is a direct consequence of topological restrictions imposed by a fiber bundle structure of the studied problem.



**Quantum Bifurcations, Fig. 19** Rearrangement of three bands into two topologically non-trivially coupled bands. Example of a model with three electronic states and vibrational structure of polyads formed by three quasi-degenerate modes. At  $\lambda = 0$  three bands have each the same number of states, namely 15. In the classical limit each initial band has rank one and trivial topology. At  $\lambda = 1$  there are only two bands. One of them has rank 2 and non-trivial first and second Chern classes. (Taken from (Faure and Zhilinskii 2002))

It is interesting to mention here the general mathematical problem of finding proper equivalence or better to say correspondence between some construction made over real numbers and their generalizations to complex numbers and quaternions. This paradigm of complexification and quaternization was discussed by Arnold (1995, 2005) on many different examples. The closest to the present subject is the example of complexification of the Wigner–Neumann non-crossing rule resulting in a quantum Hall effect (in physical terms). In fact, the mathematical basis of the quantum Hall effect is exactly the same fiber bundle construction which explains the redistribution of energy levels between bands in the above-mentioned simple quantum mechanical model.

## Multiple Resonances and Quantum State Density

Rearrangement of quantum energy states between bands is presented in the previous section as an example of a generic qualitative phenomenon occurring under variation of a control parameter. One possible realization of bands is the sequence of vibrational polyads formed by a system of resonant vibrational modes indexed by the polyad quantum number. In the classical picture this construction corresponds to the system of oscillators reduced with respect to the global action. The reduced classical phase space is in such a case the weighted projective space. In the case of particular  $1 : 1 : \dots : 1$  resonance the corresponding reduced phase space is a normal complex projective space  $CP^n$ . The specific resonance conditions impose for a quantum problem specific conditions on the numbers of quantum states in polyads. In the simplest case of harmonic oscillators with  $n_1 : n_2 : \dots : n_k$  resonance the numbers of states in polyads are given by the generating function

$$g = \frac{1}{(1 - t^{n_1})(1 - t^{n_2}) \dots (1 - t^{n_k})} = \sum_N C_N t^N, \quad (8)$$

where  $N$  is the polyad quantum number. Numbers  $C_N$  are integers for integer  $N$  values, but



they can be extended to arbitrary  $N$  values and represented in the form of a quasi-polynomial, i.e., a polynomial in  $N$  with coefficients being a periodic function whose period equals the least common multiplier of  $n_i$ ,  $i = 1, \dots, k$ . Moreover, the coefficients of the polynomial can be expressed in terms of so-called Todd polynomials which indicates the possibility of topological interpretation of such information (Michel and Zhilinskii 2001a; Zhilinskii 2001).

## Physical Applications and Generalizations

The most clearly seen physical applications of quantum bifurcations is the qualitative modification of the rotational multiplet structure under rotational excitation, i.e. under the variation of the absolute value of the angular momentum. This is related first of all with the experimental possibility to study high  $J$  multiplets (which are quite close to the classical limit but nevertheless manifest their quantum structure) and to the possibility to use symmetry arguments, which allow one to distinguish clusters of states before and after bifurcation just by counting the number of states in the cluster, which depends on the order of group of stabilizer. Nuclear rotation is another natural example of quantum rotational bifurcations (Pavlichenkov 2006). Again the interest in corresponding qualitative modifications is due to the fact that rotational bands are extremely well studied up to very high  $J$  values. But in contrast to molecular physics examples, in nuclear physics it mostly happens that only ground states (for each value of  $J$ ) are known. Thus, one speaks more often about qualitative changes of the ground state (in the absence of temperature) named quantum phase transitions (Sachdev 1999).

Internal structure of vibrational polyads is less evident for experimental verifications of quantum bifurcations, but it gives many topologically non-trivial examples of classical phase spaces on which the families of Hamiltonians depending on parameters are defined (Ezra 1996; Giacobbe et al. 2004; Joyeux et al. 2000, 2002; Kellman and Lynch 1986; Kozin et al. 2005; Sadovskii et al.

1993; Tyng and Kellman 2006; Uwano 1999; Xiao and Kellman 1990). The main difficulty here is the small number of quantum states in polyads accessible to experimental observations. But this problem is extremely interesting from the point-of-view of extrapolation of theoretical results to the region of higher energy (or higher polyad quantum numbers) which is responsible as a rule for many chemical intra-molecular processes. Certain molecules, like  $\text{CO}_2$ , or acetylene ( $\text{C}_2\text{H}_2$ ) are extremely well studied and a lot of highly accurate data exist. At the same time the qualitative understanding of the organization of excited states even in these molecules is not yet completed and new qualitative phenomena are just starting to be discovered.

Among other physically interesting systems it is necessary to mention model problems suggested to study the behavior of Bose condensates or quantum qubits (Hines et al. 2005; Hou et al. 2005; Somma et al. 2004; Wang and Kais 2004; Weyl 1952). These models have a mathematical form which is quite similar to rotational and vibrational models. At the same time their physical origin and the interpretation of results is quite different. This is not an exception. For example, the model Hamiltonian corresponding in the classical limit to a Hamiltonian function defined over  $S^2$  classical phase space is relevant to rotational dynamics, description of internal structure of vibrational polyads formed by two (quasi)degenerate modes, in particular to so-called local-normal mode transition in molecules, interaction of electromagnetic field with a two-level system, the Lipkin–Meshkov–Glick model in nuclear physics, entanglement of qubits, etc.

## Future Directions

To date many new qualitative phenomena have been suggested and observed in experimental and numerical studies due to intensive collaboration between mathematicians working in dynamical system theory, classical mechanics, complex geometry, topology, etc., and molecular physicists using qualitative mathematical tools to classify behavior of quantum systems and to extrapolate this behavior from relatively simple (low energy



regions) to more complicated ones (high energy regions). Up to now the main accent was placed on the study of the qualitative features of isolated time-independent molecular systems. Specific patterns formed by energy eigenvalues and by common eigenvalues of several mutually commuting observables were the principal subject of study. Existence of qualitatively different dynamical regimes for time-independent problems at different values of exact or approximate integrals of motion were clearly demonstrated. Many of these new qualitative features and phenomena are supposed to be generic and universal although their rigorous mathematical formulation and description is still absent.

On the other side, the analysis of the time-dependent processes should be developed. This step is essential in order to realize at the level of quantum micro-systems the transformations associated with the qualitative modifications of dynamical regimes and to control such time-dependent processes as elementary reactions, information data storage, and so on. From this global perspective the main problem of the future development is to support the adequate mathematical formulation of qualitative methods and to improve our understanding of qualitative modifications occurring in quantum micro-systems in order to use them as real micro-devices.

## Bibliography

- Arnold VI (1972) Modes and quasimodes. *Funct Anal Appl* 6:94–101
- Arnold VI (1989) *Mathematical methods of classical mechanics*. Springer, New York
- Arnold VI (1992) *Catastrophe theory*. Springer, Berlin
- Arnold VI (1995) Remarks on eigenvalues and eigenvectors of Hermitian matrices, Berry phase, adiabatic connections and quantum Hall effect. *Selecta Math New Ser* 1:1–19
- Arnold VI (2005) *Arnold's problems*. Springer, Berlin
- Aubry S, Flach S, Kladko K, Olbrich E (1996) Manifestation of classical bifurcation in the spectrum of the integrable quantum dimer. *Phys Rev Lett* 76:1607–1610
- Bolsinov AV, Fomenko AT (2004) *Integrable Hamiltonian systems. Geometry topology classifications*. Chapman and Hall/CRC, London
- Brodersen S, Zhilinskii BI (1995) Transfer of clusters between the vibrational components of  $\text{CF}_4$ . *J Mol Spectrosc* 169:1–17
- Child MS (2000) In: Jensen P, Bunker PR (eds) *Computational molecular spectroscopy*, chapter 18. Wiley Interscience, Chichester
- Child MS (2001) Quantum level structure and nonlinear classical dynamics. *J Mol Spectrosc* 210:157–165
- Child MS, Weston T, Tennyson J (1999) Quantum monodromy in the spectrum of  $\text{H}_2\text{O}$  and other systems. *Mol Phys* 96:371–379
- Colin de Verdier Y, Vũ Ngoc S (2003) Singular Bohr–Sommerfeld rules for 2D integrable systems. *Ann Ec Norm Sup* 36:1–55
- Cushman RS, Bates L (1997) *Global aspects of classical integrable systems*. Birkhäuser, Basel
- Cushman RH, Duistermaat JJ (1988) The quantum mechanical spherical pendulum. *Bull Am Math Soc* 19:475–479
- Cushman RH, Sadovskii DA (2000) Monodromy in the hydrogen atom in crossed fields. *Physica D* 142:166–196
- Dirac PAM (1982) *The principles of quantum mechanics*. Oxford University Press, Oxford
- Duistermaat JJ (1980) On global action angle coordinates. *Commun Pure Appl Math* 33:687–706
- Duistermaat JJ (1998) The monodromy in the Hamiltonian Hopf bifurcation. *Angew Z Math Phys* 49:156–161
- Efstathiou K (2004) *Metamorphoses of Hamiltonian systems with symmetry. Lecture notes in mathematics*, vol 1864. Springer, Heidelberg
- Efstathiou K, Cushman RH, Sadovskii DA (2004a) Hamiltonian Hopf bifurcation of the hydrogen atom in crossed fields. *Physica D* 194:250–274
- Efstathiou K, Joyeux M, Sadovskii DA (2004b) Global bending quantum number and the absence of monodromy in the  $\text{HCN} \leftrightarrow \text{CNH}$  molecule. *Phys Rev A* 69(3):032504-1–15
- Efstathiou K, Sadovskii DA, Zhilinskii BI (2004c) Analysis of rotation-vibration relative equilibria on the example of a tetrahedral four atom molecule. *SIAM J Dyn Syst* 3:261–351
- Efstathiou K, Cushman RH, Sadovskii DA (2007a) Fractional monodromy in the 1:–2 resonance. *Adv Math* 209:241–273
- Efstathiou K, Sadovskii DA, Zhilinskii BI (2007b) Classification of perturbations of the hydrogen atom by small static electric and magnetic fields. *Proc Roy Soc Lond A* 463:1771–1790
- Ezra GS (1996) Periodic orbit analysis of molecular vibrational spectra: spectral patterns and dynamical bifurcations in Fermi resonant systems. *J Chem Phys* 104:26–35
- Faure F, Zhilinskii BI (2000) Topological Chern indices in molecular spectra. *Phys Rev Lett* 85:960–963
- Faure F, Zhilinskii BI (2001) Topological properties of the Born–Oppenheimer approximation and implications for the exact spectrum. *Lett Math Phys* 55:219–238
- Faure F, Zhilinskii BI (2002) Topologically coupled energy bands in molecules. *Phys Lett A* 302:242–252
- Flach S, Willis CR (1998) Discrete breathers. *Phys Rep* 295:181–264

- Giacobbe A, Cushman RH, Sadovskii DA, Zhilinskii BI (2004) Monodromy of the quantum 1:1:2 resonant swing spring. *J Math Phys* 45:5076–5100
- Gilmore R (1981) Catastrophe theory for scientists and engineers. Wiley, New York
- Gilmore R, Kais S, Levine RD (1986) Quantum cusp. *Phys Rev A* 34:2442–2452
- Golubitsky M, Schaeffer DG (1984) Singularities and groups in bifurcation theory, vol 1. Springer, Berlin
- Grondin L, Sadovskii DA, Zhilinskii BI (2002) Monodromy in systems with coupled angular momenta and rearrangement of bands in quantum spectra. *Phys Rev A* 142:012105-1–15
- Guillemin V (1994) Moment maps and combinatorial invariants of Hamiltonian  $T^n$ -spaces. Birkhäuser, Basel
- Harter W (1988) Computer graphical and semiclassical approaches to molecular rotations and vibrations. *Comput Phys Rep* 8:319–394
- Hines AP, McKenzie RH, Milburn GJ (2005) Quantum entanglement and fixed-point bifurcations. *Phys Rev A* 71:042303-1–9
- Hou X-W, Chen J-H, Hu B (2005) Entanglement and bifurcation in the integrable dimer. *Phys Rev A* 71:034302-1–4
- Joyeux M, Sugny D, Tyng V, Kellman ME, Ishikawa H, Field RW, Beck C, Schinke R (2000) Semiclassical study of the isomerization states of HCP. *J Chem Phys* 112:4162–4172
- Joyeux M, Farantos SC, Schinke R (2002) Highly excited motion in molecules: saddle-node bifurcations and their fingerprints in vibrational spectra. *J Phys Chem A* 106:5407–5421
- Joyeux M, Sadovskii DA, Tennyson J (2003) Monodromy of the LiNC/NCLi molecule. *Chem Phys Lett* 382:439–442
- Joyeux M, Grebenshikov S, Bredenbeck J, Schinke R, Farantos SC (2005) Intramolecular dynamics along isomerization and dissociation pathways. In: Toda M, Komatsuzaki T, Konishi T, Berry RS, Rice SA (eds) Geometric structures of phase space in multidimensional chaos: a special volume of advances in chemical physics, part A, vol 130. Wiley, pp 267–303
- Kellman ME (1995) Algebraic models in spectroscopy. *Annu Rev Phys Chem* 46:395–422
- Kellman ME, Lynch ED (1986) Fermi resonance phase space structure from experimental spectra. *J Chem Phys* 85:7216–7223
- Kleman M (1983) Points, lines and walls. Wiley, Chichester
- Kozin IN, Sadovskii DA, Zhilinskii BI (2005) Assigning vibrational polyads using relative equilibria: application to ozone. *Spectrochim Acta A* 61:2867–2885
- Landau L, Lifschits EM (1981) Quantum mechanics, non-relativistic theory. Elsevier, Amsterdam
- Lu Z-M, Kellman ME (1997) Phase space structure of triatomic molecules. *J Chem Phys* 107:1–15
- Marsden JE, Ratiu TS (1994) Introduction to mechanics and symmetry. Springer, New York
- Mermin ND (1979) The topological theory of defects in ordered media. *Rev Mod Phys* 51:591–648
- Michel L (1980) Symmetry defects and broken symmetry, configurations, hidden symmetry. *Rev Mod Phys* 52:617–651
- Michel L, Zhilinskii BI (2001a) Symmetry, invariants, topology, vol I. Basic tools. *Phys Rep* 341:11–84
- Michel L, Zhilinskii BI (2001b) Symmetry, invariants, topology, vol III. Rydberg states of atoms and molecules. Basic group theoretical and topological analysis. *Phys Rep* 341:173–264
- Montaldi J, Roberts R, Stewart I (1988) Periodic solutions near equilibria of symmetric Hamiltonian systems. *Philos Trans R Soc Lond A* 325:237–293
- Morse M (1925) Relation between the critical points of a real function of  $n$  independent variables. *Trans Am Math Soc* 27:345–396
- Nakahara M (1990) Geometry, topology and physics. IOP Publishing, Bristol
- Nekhoroshev NN (1972) Action-angle variables and their generalizations. *Trans Moscow Math Soc* 26:180–198
- Nekhoroshev NN, Sadovskii DA, Zhilinskii BI (2006) Fractional Hamiltonian monodromy. *Ann Henri Poincaré* 7:1099–1211
- Pavlichenkov I (1993) Bifurcations in quantum rotational spectra. *Phys Rep* 226:173–279
- Pavlichenkov I (2006) Quantum bifurcations and quantum phase transitions in rotational spectra. *Phys At Nucl* 69:1008–1013
- Pavlichenkov I, Zhilinskii BI (1988) Critical phenomena in rotational spectra. *Ann Phys NY* 184:1–32
- Pavlov-Verevkin VB, Sadovskii DA, Zhilinskii BI (1988) On the dynamical meaning of the diabolic points. *Europhys Lett* 6:573–578
- Peters AD, Jaffe C, Gao J, Delos JB (1997) Quantum manifestations of bifurcations of closed orbits in the photodetachment cross section of  $H^-$  in parallel fields. *Phys Rev A* 56:345–355
- Pierre G, Sadovskii DA, Zhilinskii BI (1989) Organization of quantum bifurcations: crossover of rovibrational bands in spherical top molecules. *Europhys Lett* 10:409–414
- Sachdev S (1999) Quantum phase transitions. Cambridge University Press, Cambridge
- Sadovskii DA, Zhilinskii BI (1993) Group theoretical and topological analysis of localized vibration-rotation states. *Phys Rev A* 47(4):2653–2671
- Sadovskii DA, Zhilinskii BI (1999) Monodromy, diabolic points, and angular momentum coupling. *Phys Lett A* 256:235–244
- Sadovskii DA, Zhilinskii BI (2006) Quantum monodromy, its generalizations and molecular manifestations. *Mol Phys* 104:2595–2615
- Sadovskii DA, Zhilinskii BI (2007) Hamiltonian systems with detuned 1:1:2 resonance, manifestations of bidromy. *Ann Phys NY* 322:164–200
- Sadovskii DA, Zhilinskii BI, Champion JP, Pierre G (1990) Manifestation of bifurcations and diabolic points in molecular energy spectra. *J Chem Phys* 92:1523–1537

- Sadovskii DA, Fulton NG, Henderson JR, Tennyson J, Zhilinskii BI (1993) Nonlinear normal modes and local bending vibrations of  $H_3^+$  and  $D_3^+$ . *J Chem Phys* 99(2):906–918
- Sadovskii DA, Zhilinskii BI, Michel L (1996) Collapse of the Zeeman structure of the hydrogen atom in an external electric field. *Phys Rev A* 53:4064–4047
- Simon B (1980) The classical limit of quantum partition functions. *Commun Math Phys* 71:247–276
- Somma R, Ortiz G, Barnum H, Knill E, Viola L (2004) Nature and measure of entanglement in quantum phase transitions. *Phys Rev A* 70:042311-1–21
- Symington M (2003) Four dimensions from two in symplectic topology. In: Athens GA (ed) *Topology and geometry of manifolds. Proc symp pure math*, vol 71. AMS, Providence, pp 153–208
- Tyng V, Kellman ME (2006) Bending dynamics of acetylene: new modes born in bifurcations of normal modes. *J Phys Chem B* 119:18859–18871
- Uwano Y (1999) A quantum saddle-node bifurcation in a resonant perturbed oscillator with four parameters. *Rep Math Phys* 44:267–274
- Uzer T (1990) Zeeman effect as an asymmetric top. *Phys Rev A* 42:5787–5790
- Van der Meer JC (1985) The Hamiltonian Hopf bifurcation. *Lect notes math*, vol 1160. Springer, New York
- Vũ Ngoc S (1999) Quantum monodromy in integrable systems. *Commun Math Phys* 203:465–479
- Vũ Ngoc S (2007) Moment polytopes for symplectic manifolds with monodromy. *Adv Math* 208:909–934
- Waalkens H, Dullin HR (2001) Quantum monodromy in prolate ellipsoidal billiards. *Ann Phys NY* 295: 81–112
- Wang J, Kais S (2004) Scaling of entanglement at a quantum phase transition for a two-dimensional array of quantum dots. *Phys Rev A* 70:022301-1–4
- Weyl H (1952) *Symmetry*. Princeton University Press, Princeton
- Winnewisser M, Winnewisser B, Medvedev I, De Lucia FC, Ross SC, Bates LM (2006) The hidden kernel of molecular quasi-linearity: quantum monodromy. *J Mol Struct V* 798:1–26
- Xiao L, Kellman ME (1990) Catastrophe map classification of the generalized normal-local transition in Fermi resonance spectra. *J Chem Phys* 93:5805–5820
- Zhang W-M, Feng DH, Gilmore R (1990) Coherent states: theory and some applications. *Rev Mod Phys* 62: 867–927
- Zhilinskii BI (1996) Topological and symmetry features of intramolecular dynamics through high resolution molecular spectroscopy. *Spectrochim Acta A* 52:881–900
- Zhilinskii BI (2001) Symmetry, invariants, and topology, vol II. *Symmetry, invariants, and topology in molecular models*. *Phys Rep* 341:85–171
- Zhilinskii BI (2006) Hamiltonian monodromy as lattice defect. In: Monastyrsky M (ed) *Topology in condensed matter*. Springer series in solid state sciences, vol 150. Springer, Berlin, pp 165–186
- Zhilinskii BI, Pavlichenkov IM (1988) Critical phenomenon in the rotational spectra of water molecule. *Opt Spectrosc* 64:688–690
- Zhilinskii BI, Petrov SV (1996) Nonlocal bifurcation in the rotational dynamics of an isotope-substituted  $A_2A_2^*$  molecule. *Opt Spectrosc* 81:672–676
- Zhilinskii BI, Kozin I, Petrov S (1999) Correlation between asymmetric and spherical top: imperfect quantum bifurcations. *Spectrochim Acta A* 55:1471–1484

# Convergent Perturbative Expansion in Condensed Matter and Quantum Field Theory

Vieri Mastropietro  
Dipartimento di Matematica, Università degli Studi di Milano, Milan, Italy

## Article Outline

Glossary  
Definition of the Subject and Its Importance  
Gaussian Integrals  
Grassmann Integrals  
Perturbative Expansions  
Truncated Expectations  
Analyticity  
Conclusions  
Bibliography

## Glossary

**Condensed Matter** The study of a system of quantum nonrelativistic particles, like the conduction electrons in a metal.

**Quantum Field Theory** The theory combining special relativity and quantum mechanics.

**Bosons and Fermions** The two families of quantum particles. For instance, electrons are fermions and photons are bosons.

## Definition of the Subject and Its Importance

Quantum field theory and condensed matter have strong similarities, despite the fact that the physical phenomena which describe are rather different. In the first case, one has to deal with high-energy physics problems, where the properties of particles are described by the combination of special relativity and quantum physics. In the second case, one has to

deal with the low-energy properties of a system of an enormous number of quantum particles, whose collective behaviour determines the properties of matter; in particular, the exotic low temperature phenomena driven by quantum physics. In both cases, the physical properties are expressed by functional integrals, which can be seen as the limit of non-Gaussian integrals with variables living on a lattice; in particular, for condensed matter, one has to send the volume to infinity, while in quantum field theory also, the continuum limit has to be taken. The use of such functional integrals is the common language of condensed matter and quantum field theory, among the most successful theories in modern physics. Apart from very few cases, mainly at low dimensions, the only possibility to evaluate them is by perturbative methods.

## Gaussian Integrals

Condensed matter and quantum field theory models are typically described by an Hamiltonian or a Lagrangian expressed by the sum of two terms: the first describes the kinetic energy of the particles and the second the interaction between them. The physical properties of noninteracting particles can be exactly determined. This is not the case when the interaction is taken into account: apart from a small number of cases, typically in low dimensions, the properties of interacting systems can be studied only via perturbative methods. It should be remarked that the presence of interaction is expected in several cases to deeply modify the physical properties; this makes of course difficult to extract information from such expansions.

The physical properties are typically expressed by averages of the form

$$\langle O \rangle = \frac{\int P(d\xi) e^{-V(\xi)} O(\xi)}{\int P(d\xi) e^{-V(\xi)}} \quad (1)$$

where  $\xi \equiv \{\xi_x\}$  is the field, representing the particles, and  $x \in \Lambda$  is the particle coordinate.

The number of fields is finite, but in order to get physical information, one needs to send such number to infinity.  $V$  is a monomial, typically of degree greater than two, in the fields representing the interaction, and  $O(\xi)$  is a monomial representing the physical observable. Note that (1) essentially says that the physical properties are obtained by averaging over all the field configurations with a suitable weight.

Quantum particles can be bosons or fermions, and correspondingly  $\xi_x$  can be a Gaussian or Grassmann variables. In the first case ( $\xi_x$  is traditionally called  $\phi_x$ ),  $\phi_x \in R$  and  $P(d\phi)$  is a Gaussian measure; this means that the average of  $n$  fields can be expressed by sum of products of  $\int P(d\phi) \phi_{x_1} \phi_{x_2} = g(x_1, x_2)$ , where  $g$  is called propagator; that is the Wick rule holds

$$\int P(d\phi) \prod_{i=1}^{2n} \phi_{x_i} = \sum_{\Pi} g(x_{p_1}, x_{p_2}) \cdots g(x_{p_{2n-2}}, x_{p_{2n}}) \quad (2)$$

where  $\sum_{\Pi}$  is the sum over all the possible pairings of  $1, \dots, 2n$  (no ordering ambiguity as  $g(x_1, x_2) = g(x_2, x_1)$ ). For instance,

$$\begin{aligned} \int P(d\phi) \phi_{x_1} \phi_{x_2} \phi_{x_3} \phi_{x_4} &= g(x_1, x_2) g(x_3, x_4) \\ &\quad + g(x_1, x_3) g(x_2, x_4) \\ &\quad + g(x_1, x_4) g(x_2, x_3) \end{aligned} \quad (3)$$

The above expressions are straightforward consequences of elementary properties of Gaussian integrals; relabeling the  $\phi_x$  by a one-dimensional index and remembering that, if  $A$  is a strictly definite positive matrix of order  $n$ , then, if  $x_i \in R$ , one has

$$\begin{aligned} \int d^n x e^{-\frac{1}{2} \sum_{i,j} x_i A_{i,j} x_j} &= (2\pi)^{\frac{n}{2}} (\det A)^{-1/2} \\ \int d^n x e^{-\frac{1}{2} \sum_{i,j} x_i A_{i,j} x_j + \sum_i b_i x_i} &= (2\pi)^{\frac{n}{2}} (\det A)^{-1/2} e^{\frac{1}{2} b_i (A^{-1})_{i,j} b_j} \end{aligned} \quad (4)$$

From the second equation, the Wick rule immediately follows. A well-known example of

bosonic theory is the “ $\phi^4$  theory,” corresponding to (1) with, if  $\Lambda = a\mathbb{Z}^d \cap [-L/2, L/2]^d$  and periodic boundary conditions are imposed

$$\begin{aligned} g(x, y) &= \left(\frac{2\pi}{L}\right)^d \sum_{k=\frac{2\pi}{L}n} \frac{e^{ik(x-y)}}{\sum_{i=1}^d \frac{(1 - \cos ak_i)}{a^2} + m^2} \\ &\quad k \in (-\pi/a, \pi/a]^d \\ V &= \frac{\lambda}{4!} \sum_x \phi_x^4 \end{aligned} \quad (5)$$

## Grassmann Integrals

In the case of fermionic systems ( $\xi_x$  is traditionally called  $\psi_x$  in this case), one introduces Grassmann variables  $\psi_x^+, \psi_x^-$ , such that  $\{\psi_x^+, \psi_x^-\} = 0$ . A Grassmann integral is defined by the following rule

$$\int d\psi_x = 0 \quad \int \psi_x d\psi_x = 0 \quad (6)$$

$P(d\psi)$  is a Grassmann Gaussian integration defined by the anticommutative Wick rule

$$\begin{aligned} \int P(d\psi) \prod_{i=1}^n \psi_{x_i}^- \psi_{y_i}^+ &= \sum_{\Pi} g(x_{p_1}, x_{p_2}) \cdots g(x_{p_{2n-2}}, x_{p_{2n}}) \\ &= \sum_P (-1)^{\varepsilon_P} \prod_{i=1}^n g(x_i, y_{p_i}) \\ &= \det G \end{aligned} \quad (7)$$

where  $P$  is the set of all possible permutations of  $1, \dots, n$ , and  $\varepsilon_P$  is the sign of the permutation; moreover,

$$G = \begin{pmatrix} g(x_1, y_1) & g(x_1, y_2) & \cdots & g(x_1, y_n) \\ g(x_2, y_1) & g(x_2, y_2) & \cdots & g(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ g(x_n, y_1) & g(x_n, y_2) & \cdots & g(x_n, y_n) \end{pmatrix} \quad (8)$$

For instance,

$$\int P(d\psi) \psi_{x_1}^- \psi_{x_2}^- \psi_{y_1}^+ \psi_{y_2}^+ = g(x_1, y_1) g(x_2, y_2) - g(x_1, y_2) g(x_2, y_1) \quad (9)$$

Note the presence of relative signs in the permutations, a fact producing, as we will see, dramatic

differences in the convergence properties of the series expansions.

As an example of fermionic theory, we can consider the Hubbard model, defined by (1) with  $\Lambda = (0, \beta) \cup \overline{\Lambda}$  and  $\overline{\Lambda} = \mathbb{Z}^d \cap [-L/2, L/2]^d$ , with  $\psi_{x,\sigma}^\pm$ ,  $\sigma = \pm$ ,  $x = x_0, \dots, x_d$  with antiperiodic boundary conditions in  $x_0$  and periodic in  $x_1, \dots, x_d$ , and

$$g(x, y) = \left(\frac{2\pi}{\beta}\right) \sum_{k_0 = \frac{2\pi}{\beta}(n + \frac{1}{2})} \left(\frac{2\pi}{L}\right)^d \sum_{\substack{k = \frac{2\pi}{L}n \\ k \in (-\pi, \pi]^d}} \frac{e^{ik(x-y)}}{-ik_0 + \sum_{i=1}^d \cos k_i - \mu} \quad (10)$$

$$V = \lambda \sum_x \psi_{x,+}^+ \psi_{x,+}^- \psi_{x,-}^+ \psi_{x,-}^-$$

It is remarkable that the same general expression (1) holds for condensed matter and quantum field theory; such fields describe indeed radically different phenomena from the physical point of view. In the case of condensed matter, one describes a system of nonrelativistic quantum particles; a typical example is a system of conduction electrons in a metal interacting via electric forces, like the Hubbard model introduced above. In this case,  $x_0$  represents the (imaginary) time,  $\beta^{-1}$  is the temperature, and  $L$  is the side of the box containing the fermions. In physical applications, the length of the box is immensely larger than the reticular step (assumed equal to 1 in (10)), and it can be considered to all purposes infinity; that is the observable has to be computed in the thermodynamic limit  $L \rightarrow \infty$ . If one is interested in the low-temperature properties, which is the regime where quantum effects appear more dramatically, also the  $\beta \rightarrow \infty$  has to be taken. On the contrary, the lattice step has to be kept finite; actually in condensed matter, it has a physical meaning, i.e., the distance of the ions forming the crystal in which the conduction electrons move. Quantum field theory describe instead the interaction of relativistic quantum particles; the lattice step is in this case a mere regularization, or an ultraviolet cut-off, which should be

removed at the end  $a \rightarrow 0$ ; an example is the  $\phi^4$  model (5). The limit  $L \rightarrow \infty$  is called infrared limit, while  $a \rightarrow 0$  is the ultraviolet limit.

## Perturbative Expansions

In order to compute (1), a possibility is expanding in Taylor series the exponential, that is

$$Z \equiv \int P(d\xi) e^{-V(\xi)} = \int P(d\xi) \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} V^n \quad (11)$$

Assuming that we can exchange the sum with the integral we get, if  $V$  has the form  $V = \lambda \sum_x M(\xi)$ , with  $M$  a monomial in  $\xi$

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int P(d\xi) V^n \\ &= \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} \sum_{x_1, \dots, x_n} \mathcal{E}(M(\xi_{x_1}) \dots M(\xi_{x_n})) \\ &\equiv \sum_{n=0}^{\infty} \lambda^n c_n \end{aligned} \quad (12)$$

where we use the notation



$$\mathcal{E}(M(\xi_{x_1}) \dots M(\xi_{x_n})) = \int P(d\xi) M(\xi_{x_1}) \dots M(\xi_{x_n}) \quad (13)$$

and  $\mathcal{E}$  is the simple expectation, which can be computed by the Wick rule.

Note that (12) is a power series in  $\lambda$ , and a first question concerns its convergence at finite  $L$  and  $a$ . The situation is different between bosons and fermions. In the latter case, due to the fact that  $(\psi_x^+)^2 = 0$  or  $(\psi_x^-)^2 = 0$  due to the anti-commutation rules, one has that the series is just a finite sum. This is evident, for instance, in the example (10); the number of points in  $\Lambda$  is  $L^d$  and each  $V$  involves two coordinates, so that for  $n \geq N$  with  $N = O(L^d)$ , there are necessarily two fields associated to the same coordinate, hence  $c_n = 0$ . Physically this fact is connected with Pauli principle. Different is the case of bosons, where the series consists indeed of infinite terms and no analyticity is expected. The difference between Gaussian and Grassmann integrals is already evident in the  $d = 0$  dimensional case where

$$Z = \int_{-\infty}^{\infty} dx e^{-x^2 - \lambda x^4} \quad (14)$$

and the associated series is

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} dx e^{-x^2} (-\lambda x)^n \\ &= \sum_{n=0}^{\infty} \sqrt{\pi} (-1)^2 \lambda^n \frac{(4n-1)!!}{2^n n!} \equiv c_n \lambda^n \end{aligned} \quad (15)$$

As  $c_n$  grows as  $O(n!)$ , the power series cannot be convergent; this was a priori evident as the integral cannot be analytic around  $\lambda = 0$  as is divergent for negative  $\lambda$ . The situation is, however, radically different for Grassmann integral where

$$\begin{aligned} Z &= \int d\psi_1^+ d\psi_1^- d\psi_2^+ d\psi_2^- e^{-\psi_1^+ \psi_1^- - \psi_2^+ \psi_2^- - \lambda \psi_1^+ \psi_1^- \psi_2^+ \psi_2^-} \\ &= 1 - \lambda \end{aligned} \quad (16)$$

Even in the case of fermions, where (12) is a finite sum for finite dimensional integrals, the use

of (12) is problematic. In physical applications, one is interested in taking the infinite volume limit  $L \rightarrow \infty$ . The coefficient  $c_n$  in (12) is expressed by a sum of terms, coming from the computation of the expectations by the Wick rule. Among such terms, one can easily identify terms  $O(L^{dn})$  that is growing with the volume with a power given by the perturbative order. This means that higher orders are larger than lower ones for large volumes. It is sometime convenient to represent the different contributions to  $c_n$  by graphical elements, called Feynman graphs. Such graphs can be connected or disconnected, and it turns out that the power of  $L^d$  correspond to the number of disconnected pieces plus one.

## Truncated Expectations

There is a clear physical motivation in getting the growth  $O(L^{dn})$  at order  $n$ ; in the condensed matter interpretation,  $\log Z$  is the free energy which is an extensive quantity, that is proportional to the volume. It is therefore not convenient to expand  $Z$  in powers of the coupling  $\lambda$ ;  $Z$  behaves as the exponential of the volume so that the coefficients of the series in  $\lambda$  are  $O(L^{dn})$ . It is instead much more convenient to expand  $\log Z$ .

In order to define an expansion for such a quantity, one has to introduce the truncated expectations, defined by

$$\mathcal{E}^T(M, n) = \frac{\partial^n}{\partial \lambda^n} \log \int P(d\xi) e^{\lambda M(\xi)} \Big|_0 \quad (17)$$

For instance,

$$\mathcal{E}^T(M, 2) = \mathcal{E}(M^2) - (\mathcal{E}(M))^2 \quad (18)$$

Therefore  $\log Z$  can be (formally) written as

$$\begin{aligned} \log Z &= \int P(d\xi) e^{-\lambda V(\xi)} \Big|_{\lambda=1} = \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{1}{n!} \frac{\partial^n}{\partial \lambda^n} \log \int P(d\xi) e^{\lambda V(\xi)} \Big|_0 \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{1}{n!} \mathcal{E}^T(V; n) \end{aligned} \quad (19)$$

and

$$\mathcal{E}^T(V; n) = \sum_{x_1, \dots, x_n} \lambda^n \mathcal{E}^T(M(\xi_{x_1}); \dots; M(\xi_{x_n})) \quad (20)$$

It turns out that the  $n$ -th order is  $O(L^d)$ , and it can be expressed in terms of Feynman diagrams which are all connected. Note that  $Z$  was the denominator of (1), but the representation as sum of truncated expectations holds also for the average; it is sufficient to define a generating function given by

$$e^{W(\varepsilon)} = \int P(d\xi) e^{-V(\xi) - \varepsilon O(\xi)} \quad (21)$$

so that the average (1) is given by  $\frac{\partial W}{\partial \varepsilon}|_{\varepsilon=0}$ , and  $W(\varepsilon)$  can be again represented as sum of truncated expectations.

We discuss now the properties of the expansion for  $\log Z$ , focusing in particular on the two examples of  $\phi^4$  and Hubbard model; but the consideration are rather general. As we noticed at order  $n$ , the perturbative expansion is given by the sum of terms represented by connected Feynman diagrams. They are obtained representing each  $V$  as a vertex expressed by four (for quartic interactions) half lines coming from a point  $x$ ; then a graph contributing at order  $n$  is obtained considering  $n$  vertices and connecting all the half lines so that the graph is connected. To each graph is associated a value given by the sum over all the coordinates  $x_1, \dots, x_n$  of the product of propagators  $g(x_i, x_j)$  with  $x_i, x_j$  the edge of lines in the graph. Once more, it is important to stress that such graphs are just a way to represent pictorially the terms coming from the Wick rule. As all the points are connected, we can use a tree of propagators to perform the sum over the coordinates, and we can bound by their sum the remaining propagators. Therefore each Feynman graph is bounded by

$$L^d \max(|g|_1, |g|_\infty)^{2n} \frac{1}{n!} |\lambda|^n \quad (22)$$

where  $|g|_1$  and  $|g|_\infty$  are respectively the  $L_1$  and  $L_\infty$  norm of the propagator, and we have used the propagator  $g$  to perform the sums over all coordinates except one. The graphs divided by the volume are finite if  $|g|_1, |g|_\infty$  are finite; even if this is

the case, it is not sufficient to achieve convergence. Indeed in the case of quartic interactions, the number of graphs contributing to order  $n$  is  $O(n!^2)$ , so that at the end the following bound is obtained:

$$\frac{1}{L^d} \left| \log \int P(d\psi) e^{-V} \right| \leq \sum_{n=0}^{\infty} n! \max(|g|_1, |g|_\infty)^{2n} |\lambda|^n \quad (23)$$

The bound is the same in the bosonic or fermionic case, but there is a deep difference between the two possibilities. In the case of bosonic models, the dependence on  $n$  cannot be improved; the order of the perturbative expansion really grows at least as a factorial, as a consequence of the lack of analyticity in the coupling at the origin.

In contrast, the bound for fermionic models is pessimistic as it ignores the possible compensations between graphs, coming from the anticommutativity of Grassmann variables.

## Analyticity

Technically one can exploit the fact that the anticommutative Wick rule is expressed by a determinant and use the Gram-Hadamard bound. We recall that, given a matrix of the form  $M_{\alpha,\beta} = (f_\alpha, g_\beta)$  where  $\alpha, \beta = 1, \dots, n$ ,  $f_\alpha, g_\beta$  are vectors in an Hilbert space, if  $(\cdot, \cdot)$  is the scalar product, then

$$|\det M| \leq \prod_{\alpha=1}^n \|f_\alpha\| \|g_\alpha\| \quad (24)$$

where  $\|f\|^2 = (f, f)$ . We apply to the fermionic expectations using that  $\mathcal{E}(\psi \dots \psi) = \det G$  with  $G_{\alpha,\beta} = g(x_\alpha, y_\beta)$ . Setting  $f_\alpha$  as  $g(x_\alpha; -)$  and  $g_\alpha = \delta_{\alpha,-}$  one gets, for quartic  $V$

$$\begin{aligned} |\mathcal{E}(V, \dots V)| &\leq L^{dn} \prod_{\alpha=1}^{2n} \sqrt{\sum_{\beta} |g(x_\alpha, y_\beta)|^2} \\ &\leq L^{dn} (\sqrt{2n})^{2n} |g|_\infty^{2n} \end{aligned} \quad (25)$$

Therefore

$$\left| \int P(d\psi) e^{-V} \right| \leq \sum_{n=0}^{\infty} C^n (|g|_{\infty})^{2n} (L^d)^n |\lambda|^n \quad (26)$$

There is no  $n!$  in the above bound but at the cost of a factor  $L^{dn}$ .

Remarkably one can take advantage of Gram-bounds even in the truncated expectations. In most physical applications, the propagator admits the following Fourier representations

$$g(x, y) = \frac{1}{L^d} \sum_k e^{ik(x-y)} \widehat{g}(k) \quad (27)$$

hence it can be written in the form of a scalar product  $g(x, y) = \sum_z \bar{A}_x(z) B_y(z) = (A, B)$  with

$$\begin{aligned} A_x(z) &= \frac{1}{L^d} \sum_k e^{ik(x-z)} |\widehat{g}(k)|^2 \\ B_x(z) &= \frac{1}{L^d} \sum_k e^{ik(x-z)} 1/\widehat{g}(k)^* \end{aligned} \quad (28)$$

If  $M(\psi_x)$  is a monomial quartic in  $\psi$  ( $x$  are the coordinates), the following bound holds

$$\begin{aligned} |\mathcal{E}^T(M(\psi_{x_1}); \dots; (M(\psi_{x_n}))| \\ \leq C^n (|A|_{\infty} |B|_{\infty})^n \sum_T \prod_l g(x_l, y_l) \end{aligned} \quad (29)$$

where  $T$  is a set of lines  $l$  connecting the points  $x_1, \dots, x_n$  and  $x_l, y_l$  are the coordinates associated to each line. The bound is proved extracting from the expectation a tree of propagators connecting the points, bounding all the other fields by the Gram-Hadamard inequality for determinants. By (20) one gets

$$\begin{aligned} \frac{1}{L^d} |\log Z| &\leq \sum_{n=0}^{\infty} \frac{C^n}{n!} \sum_{x_1, \dots, x_n} \lambda^n \mathcal{E}^T(M(\zeta_{x_1}; \dots; \zeta_{x_n})) \\ &\leq \sum_{n=0}^{\infty} \frac{1}{n!} C^n (|A|_{\infty} |B|_{\infty})^n \sum_{x_1, \dots, x_n} \sum_T \prod_l g(x_l, y_l) \end{aligned} \quad (30)$$

The key point is that, by Cayley formula, the number of trees is bounded by  $n!C^n$  so that

$$\frac{1}{L^d} |\log Z| \leq \sum_{n=0}^{\infty} C^n (|A|_{\infty} |B|_{\infty})^n (|g|_1)^n \quad (31)$$

implying convergence of the power series for  $|A|_{\infty} |B|_{\infty}, |g|_{\infty}$  finite.

The above convergence result implies in particular analyticity in the Hubbard model. Due to the fact that  $|k_0| \geq \beta^{-1}$ , one gets a bound for the propagator given by  $|g|_1 \leq C\beta^{-1}$  so that one has convergence for temperatures  $T \geq C|\lambda|^{-1}$ .

## Conclusions

Of course in most physical applications either  $|g|_1$  and  $|g|_{\infty}$  are not finite uniformly in  $L$  and  $a$ . In condensed matter problems due to the finite lattice step  $|g|_{\infty} < \infty$  but  $|g|_1$  is unbounded at zero temperature in the infinite volume limit. In quantum field theory models, one is dealing with the limit  $a \rightarrow 0$ , so also  $|g|_1$  is unbounded. Even in such cases, it is in certain cases possible to have analyticity but the proof is much more intricate and requires partial resummations in the expansion. Analyticity at zero temperature in the infinite volume limit has been proven in condensed matter models like one-dimensional fermionic systems on a lattice or the bidimensional Hubbard model on the honeycomb lattice. At smaller temperatures, one can achieve convergence in the Hubbard model on the square lattice up to exponentially small temperature (at lower temperature, the breakdown of series is expected due to quantum phenomena like superconductivity). Similarly, quantum field theory models in  $d = 1 + 1$  dimensions, like Thirring or Yukawa models, have been constructed and have proven to be analytic in the continuum and infinite volume limit. The resummation of the series is done using a multiscale method based on renormalization group analysis, and the identity between the resummed and original expansion is established using analytic continuation. There are also cases of asymptotically free fermionic theories for which Borel summability can be proven; they are still in the realm of perturbative methods, implemented via renormalization group techniques. In the case of bosonic models like  $\phi^4$ , there is no convergence of

the series expansions; however, the lack of analyticity is due to the contribution of large fields, and in several cases, their contributions are small so that perturbative expansions give correct information, as rigorously proven in several  $\phi^4$  models. It should anyway be remarked that the use of perturbative methods in physics goes of course much beyond the cases in which it can be mathematically justified, and it is an invaluable source of physical information and mathematical problems.

## Bibliography

- Bauerschmidt R, Brydges D, Slade G (2019) Introduction to a renormalisation group method. Springer, Germany
- Benfatto G, Gallavotti G (1995) Renormalization group. Princeton University Press, Princeton
- Mastropietro V (2008) Non perturbative Renormalization. World Scientific, Singapore
- Rivasseau V (1991) From perturbative to constructive renormalization. Princeton University Press, Princeton
- Salmhofer M (1999) Renormalization. Springer, Germany

# Correlation Corrections as a Perturbation to the Quasi-free Approximation in Many-Body Quantum Systems

Niels Benedikter and Chiara Boccato  
Dipartimento di Matematica ‘Federigo Enriques’,  
Università degli Studi di Milano, Milan, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
The Framework of Quantum Mechanics  
Many-Body Quantum Mechanics  
Quantities of Interest  
Article Roadmap  
Bosons: Low-Density Scaling Limit (Gross–Pitaevskii Limit)  
Fermions: High-Density Scaling Limit (Mean-Field/Semiclassical Limit)  
Bogoliubov Transformations and Quasi-free States  
Quasi-free Approximations  
Correlation Corrections to the Gross–Pitaevskii Approximation  
Correlation Corrections to the Hartree–Fock Approximation  
Future Directions  
References

## Glossary

**Bogoliubov theory** Quadratic approximation for the many-boson Hamiltonian when acting on states exhibiting Bose–Einstein condensation (in particular states close to the ground state).  
**Bogoliubov transformation/map** Transformation of operators on Fock space mapping creation

operators into a linear combination of creation and annihilation operators such that the CCR/CAR are satisfied also by the transformed operators. Interpreted as quasiparticles.

**Bose–Einstein condensate** A state of a bosonic system in which most particles occupy the same one-particle wave function  $\varphi \in L^2(\mathbb{R}^3)$ . More precisely, a state  $\psi_N \in L^2_{\text{symm}}(\mathbb{R}^{3N})$  close to a tensor product  $\varphi^{\otimes n}$  the sense that its one-particle reduced density matrix converges to the projection  $|\varphi\rangle\langle\varphi|$  as particle number  $N \rightarrow \infty$ . Experimentally realized in atomic gases at temperatures close to absolute zero.

**Boson** Particle with the property that a system of identical particles (i.e., having same mass, same charge, same interactions) has wave function that is symmetric under permutation of the particle coordinates  $x_i \in \mathbb{R}^3$ . It is a principle of physics that all particles can be classified in bosons and fermions.

**Canonical (anti)commutation relations, CCR/CAR** The bosonic creation and annihilation operators (i.e., defined on the symmetric Fock space) satisfy the CCR:  $[a(f), a(g)] = 0$  and  $[a(f), a^*(g)] = \langle f, g \rangle_{L^2(\mathbb{R}^3)}$ , where  $[A, B] := AB - BA$  is the commutator of two operators  $A, B$ . In the fermionic case, the commutator is replaced by the anticommutator  $\{A, B\} := AB + BA$  and the relations are called CAR.

**Correlation energy** Difference between the ground state energy and the lowest energy achievable with quasi-free states.

**Creation and annihilation operators** Creation operators take the  $n$ -particle element of a Fock space vector and add an additional particle to it (using the tensor product and symmetrizing/antisymmetrizing for bosonic/fermionic particles, respectively), taking it to the  $n + 1$ -particle sector. They are adjoint to annihilation operators, which integrate out one particle. In the bosonic case, they satisfy the CCR, and in the fermionic case, the CAR, which provide a convenient way of computing with many-body vectors and Hamiltonians.

**Fermion** A particle with the property that a system of identical particles has wave function that is antisymmetric under permutation of particles. It is a physical principle that particles can be classified in bosons and fermions.

**Fermi ball** The ground state of  $N$  noninteracting fermions on the three-dimensional torus is given by the Slater determinant of the  $N$  plane waves  $f_k(x) = e^{ik \cdot x}$  of lowest kinetic energy, i.e., with momenta  $k \in 2\pi\mathbb{Z}^3$  forming a ball in Fourier space centered on the origin, with radius proportional to  $N^{1/3}$ , called the Fermi ball.

**Fock space** The direct sum  $\mathcal{F}_{s/a}$  (over particle number  $m$ ) of all  $m$ -particle Hilbert spaces of a given type of system. The procedure of embedding an  $N$ -particle Hilbert space  $L^2_{s/a}(\mathbb{R}^{3N})$  in Fock space and writing the Hamiltonian with creation and annihilation operators is known as second quantization.

**Gross–Pitaevskii approximation** Quasi-free approximation for the ground state energy of bosonic systems interacting through a potential with scattering length proportional to  $N^{-1}$ .

**Ground state (energy)** The ground state energy is the infimum of the spectrum of the Hamiltonian,  $E_N = \inf \sigma(H_N)$ , or equivalently given by the variational problem  $E_N = \inf_{\psi} \langle \psi, H_N \psi \rangle$ , where the latter infimum is over all normalized vectors  $\psi$  in the Hilbert space on which  $H_N$  acts. Under general assumptions on the interaction potential (Reed and Simon 2012),  $E_N$  is an eigenvalue of  $H_N$ ; all normalized associated eigenvectors are called ground states.

**Hamiltonian** A self-adjoint linear operator on a Hilbert space of wave functions (such as  $L^2(\mathbb{R}^3)$  for a one-particle system, or  $L^2_{s/a}(\mathbb{R}^{3N})$  or  $\mathcal{F}_{s/a}$  for a many-body system). Its spectrum corresponds to the energies attainable by the system. The Hamiltonian generates the dynamics.

**Hartree–Fock theory** The quasi-free approximation for fermionic systems without Cooper pairing, obtained by restricting the Hilbert space  $L^2_a(\mathbb{R}^{3N})$  to its submanifold of Slater determinants.

**Lee–Huang–Yang formula** Formula for the correlation energy of a dilute Bose–Einstein condensate, proposed by Lee et al. (1957).

**Pairing density** Also called the Cooper pair wave function, for  $\psi \in \mathcal{F}_{s/a}$  defined using annihilation operators as  $\alpha(x, y) = \langle \psi, a_x a_y \psi \rangle$ . A nonvanishing  $\alpha \neq 0$  may carry a superconducting current.

**Particle-hole transformation** A special case of a fermionic Bogoliubov transformation with the property that the transformed vacuum is a Slater determinant.

**Perturbation theory** By “perturbation theory,” we refer to conventional perturbation theory as a power series expansion in the small coupling constant  $\lambda$ , typically in the form of Rayleigh–Schrödinger or diagrammatic perturbation theory.

**(Quantum) Correlations** The Hilbert space of a system of  $N$  distinguishable particles is given by an  $N$ -fold tensor product, e.g.,  $L^2(\mathbb{R}^{3N}) \simeq L^2(\mathbb{R}^3)^{\otimes N}$ . If a vector  $\psi \in L^2(\mathbb{R}^3)^{\otimes N}$  is an elementary tensor, i.e.,  $\psi = \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_N$ , it is called an uncorrelated state; if it cannot be written as an elementary tensor, it is called a correlated (or entangled) state. For  $N$  indistinguishable bosons, i.e., in the subspace of symmetrized tensor products, uncorrelated states are of the form  $\psi = \varphi^{\otimes N}$ . In the fermionic case, all states are correlated due to antisymmetrization; we say that Slater determinants (being antisymmetrized elementary tensors) have the trivial or minimal amount of correlations. Quasi-free states contain correlations; in this chapter, we discuss the role of correlations that fall outside the set of quasi-free states.

**Quasi-free state** Vector in Fock space that can be written as a Bogoliubov-transformed vacuum satisfies the Wick theorem and is thus completely determined by its one-particle reduced density matrix  $\gamma^{(1)}$  and its pairing density  $\alpha$ .

**Random phase approximation (RPA)** The RPA was introduced by Bohm and Pines (1953) to compute the correlation energy of fermions with Coulomb interaction; they argued that certain terms in a transformation of the Hamiltonian are given by a phase depending on the position of all electrons, which are randomly distributed, therefore cancelling out. The RPA was reformulated as an effective bosonic theory (Sawada 1957;



Sawada et al. 1957) and as a partial resummation of the conventional perturbation series (Gell-Mann and Brueckner 1957).

**Reduced density matrix** Given an  $N$ -particle wave function  $\psi \in L^2_{s/a}(\mathbb{R}^{3N})$ , the  $k$ -particle reduced density matrix  $\gamma^{(k)}$  is an operator on  $L^2(\mathbb{R}^{3k})$  defined by the partial trace over  $N - k$  particles of the projection on  $\psi$ , i.e.,  $\gamma^{(k)} := \text{tr}_{N-k} |\psi\rangle\langle\psi|$ . For  $\psi \in \mathcal{F}_{s/a}$ , in terms of creation and annihilation operators, one has (in integral kernel notation)  $\gamma^{(1)}(x; y) = \langle \psi, a_y^* a_x \psi \rangle$ .

**Scaling limit** A choice of the system parameters (coupling constants, volume, interaction potential) as functions of particle number  $N$ , modelling physical situations (high density, low density, ...) as the asymptotics for  $N \rightarrow \infty$  are considered.

**Scattering equation** Describes zero-energy scattering processes among two particles. It is the two-body time-independent Schrödinger equation  $E_2\psi = H_2\psi$ , with  $E_2 = 0$ , expressed in the relative coordinates to the center of mass, i.e.,

$$\left(-\Delta + \frac{1}{2}V(x)\right)f(x) = 0 \quad (1)$$

with the boundary condition

$$f(x) \rightarrow 1 \text{ as } |x| \rightarrow \infty.$$

**Scattering length** Parameter associated to an interaction potential  $V$  giving a combined measure of its strength and its range. It is defined as

$$8\pi a = \int V(x)f(x)dx, \quad (2)$$

where  $f$  is the solution of the scattering equation.

**Second quantization** Representation of an  $N$ -body system on Fock space using creation and annihilation operators.

**Slater determinant** An antisymmetrized tensor product in  $L^2_a(\mathbb{R}^{3N})$ , i.e., a wave function of the form  $\psi(x_1, \dots, x_N) = (N!)^{-1/2} \det(\varphi_i(x_j))$  where  $\varphi_i \in L^2(\mathbb{R}^3)$ . Fermionic quasi-free

states with pairing density  $\alpha = 0$  are Slater determinants.

**State** We identify the state of a quantum system with vectors in its Hilbert space.

**Vacuum** The vector  $\Omega = (1, 0, 0, \dots)$  in the bosonic or fermionic Fock space  $\mathcal{F}_{s/a}$ .

**Weyl operator, coherent state** Given any  $\varphi \in L^2(\mathbb{R}^3)$ , the Weyl operator is defined as,  $W(\varphi) = \exp(a^*(\varphi) - a(\varphi)) : \mathcal{F}_{s/a} \rightarrow \mathcal{F}_{s/a}$ . By applying it to the vacuum, one obtains coherent states, which are eigenvectors of the annihilation operators and describe a Bose–Einstein condensate.

**Wick theorem** Expresses the expectation value of any product of creation and annihilation operators as a sum over products of the one-particle reduced density matrix  $\gamma^{(1)}$  and the pairing density  $\alpha$ . Holds on quasi-free states.

## Definition of the Subject

Many-body quantum mechanics is the physical theory describing much of our world. Its fundamental equation, the Schrödinger equation, allows in principle for the computation of all the physical properties of a system, e.g., the time evolution and the characteristic energy levels. In practice, however, the Schrödinger equation cannot be solved exactly anymore if the number of particles exceeds about three; numerical methods are hindered by the exponential growth of dimension, and perturbation theory often runs into convergence problems (Perturbation theory for bosonic systems cannot provide a convergent series expansion. This is easiest seen from the functional integral representation, which, just like the ordinary integral  $\int_{\mathbb{R}} e^{-x^2 + \lambda x^4} dx$ , can be expanded into an asymptotic series w. r. t.  $\lambda$ ; however, the series has vanishing radius of convergence because the integral is  $+\infty$  for every  $\lambda < 0$ ).

The analysis of scaling limits provides an alternative expansion (compared to conventional perturbation theory in powers of the interaction potential) of the observable physical properties. Instead of a power series in a coupling parameter, one considers the particle number  $N \rightarrow \infty$  and lets all further parameters (system volume, density, coupling parameter) depend on it. One obtains

expansions with respect to  $N$ , in which the leading order is given by a quasi-free theory. The next-to-leading order depends on interaction-induced quantum correlations (nontrivial linear combinations in the Hilbert space of the system); typically it involves all orders of the interaction potential, formally corresponding to a partial resummation of the conventional perturbation theory. We may interpret this next-to-leading order as arising from a perturbation of the quasi-free theory by quantum correlations. We illustrate this point of view with two important examples: the Gross–Pitaevskii scaling limit for bosonic systems and a mean-field/semiclassical scaling limit for fermionic systems.

## Introduction

Through many-body quantum mechanics, we describe a huge variety of systems: such diverse physical behavior as that of metals, semiconductors, organic molecules, or neutron stars can all be modelled using the Schrödinger equation. In principle, the Schrödinger equation allows for the exact computation of stationary energy levels of a physical system (and transitions between stationary energy levels are maybe the quantum-mechanical properties most accessible to experiment). In practice, however, the applicability of the Schrödinger equation is limited: exact solutions generally become impossible for more than two particles. Numerical methods are hindered by the enormous number of degrees of freedom (which grows exponentially with the number of particles, and the number of particles itself may already reach  $10^{24}$  in our macroscopic world). Perturbation theory (typically formulated using Feynman diagrams) is very hard to control in a mathematically rigorous fashion as to even obtain finite results. We present here a recent alternative approach to the mathematically rigorous analysis of quantum many-body systems, not plagued by the mentioned problems. This approach is the analysis of scaling limits, where the particle number  $N \rightarrow \infty$  is used as the parameter for obtaining an expansion of the observable quantities. The different orders in the obtained expansions are

given through certain effective theories, which are numerically or analytically more accessible than the general problem.

We start by reviewing the foundations of many-body quantum mechanics and then introduce the most important scaling limits that have been extensively studied in recent years. Afterward, we discuss the main theoretical tools: second quantization, Bogoliubov theory, and quasi-free approximations. Finally, we come to the main topic of the present note and describe key results of the last years where the expansion of scaling limits has been pushed to go beyond quasi-free states in the description of quantum correlations.

## The Framework of Quantum Mechanics

In quantum mechanics, the state of a physical system is described by a vector  $\psi$  in a Hilbert space  $\mathcal{H}$ . In the simplest case, a single particle moving in three-dimensional space, one has  $\mathcal{H} = L^2(\mathbb{R}^3)$ . The vector  $\psi \in \mathcal{H}$  is then also called a one-particle wave function, and  $\int_{\Omega} |\psi(x)|^2 dx$  is interpreted as the probability to find the particle in region  $\Omega \subset \mathbb{R}^3$ . In particular, the total probability to have the particle anywhere in the system is  $\|\psi\|^2 := \int_{\mathbb{R}^3} |\psi(x)|^2 dx = 1$ .

The time evolution of a quantum-mechanical system is described by the Schrödinger equation: given a linear self-adjoint operator  $H$ , called the Hamiltonian, given the wave function  $\psi \in L^2(\mathbb{R}^3)$  at time  $t = 0$ , the state at time  $t > 0$  is obtained by solving the initial value problem (where  $\hbar$  is the Planck constant)

$$i\hbar \partial_t \psi_t = H \psi_t, \quad \psi_0 = \psi.$$

A typical Hamiltonian operator (or *Hamiltonian*) for a one-particle problem is of the form

$$H = -\hbar^2 \Delta + V_{\text{ext}}(x),$$

where  $V_{\text{ext}}(x)$  is a multiplication operator describing an external potential. If  $V_{\text{ext}}(x) = -|x|^{-1}$  (the attractive Coulomb potential), this describes the hydrogen atom.

We are interested in more complex systems: instead of a single particle interacting with an external, fixed, potential, we consider the physically more realistic case in which many particles all move and interact with each other. This brings us to many-body quantum mechanics.

## Many-Body Quantum Mechanics

We consider a special case of many-body quantum mechanics: that of  $N$  particles moving in  $\mathbb{R}^3$  and each interacting pairwise with all other particles. This is in fact not a big restriction: systems such as metals, cold rubidium gases, superfluids, semiconductors, even superconductors fall in this category. As a single particle is described by a wave function  $\psi \in L^2(\mathbb{R}^3)$ , one possible state for a many-body system is a wave function

$$\begin{aligned}\psi(x_1, x_2, \dots, x_N) &= \varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_N(x_N), \\ \varphi_i &\in L^2(\mathbb{R}^3), \quad x_i \in \mathbb{R}^3.\end{aligned}$$

However, as quantum mechanics is a linear theory, also linear combinations (known as entangled states) of such products are possible states of the system; this means that the  $N$ -body system is described by vectors in the tensor product Hilbert space  $L^2(\mathbb{R}^3)^{\otimes N} \simeq L^2(\mathbb{R}^{3N})$ .

The prototypical Hamilton operator is of the form

$$\begin{aligned}H_N &= \sum_{i=1}^N (-\hbar^2 \Delta_{x_i} + V_{\text{ext}}(x_i)) \\ &\quad + \lambda \sum_{1 \leq i < j \leq N} V(x_i - x_j).\end{aligned}\tag{3}$$

The Laplace operators  $\Delta_{x_i}$  model the kinetic energy of particles; the interaction potential  $V: \mathbb{R}^3 \rightarrow \mathbb{R}$  is a function acting as a multiplication operator, here with respect to all pairs of particles.

The parameter  $\lambda \in \mathbb{R}$  is a physical coupling constant.

There is an additional important aspect in the description of systems of indistinguishable (same mass, same charge, same spin, etc.) particles: as one of the fundamental principles of physics, all particles may be classified as bosons or fermions. *Bosons* are particles described by a symmetric wave function  $\psi \in L^2(\mathbb{R}^{3N})$ , i.e., invariant under all permutations of the particle coordinates,

$$\begin{aligned}\psi(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}, \dots) \\ = \psi(x_1, x_2, x_3, \dots) \quad \forall \sigma \in S_N.\end{aligned}$$

*Fermions* are particles described by an antisymmetric wave function  $\psi \in L^2(\mathbb{R}^{3N})$ , i.e., acquiring a sign for each transposition of the particle coordinates,

$$\begin{aligned}\psi(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}, \dots) \\ = \text{sgn}(\sigma) \psi(x_1, x_2, x_3, \dots) \quad \forall \sigma \in S_N.\end{aligned}$$

We write  $L_s^2(\mathbb{R}^{3N})$  for the symmetric (bosonic) and  $L_a^2(\mathbb{R}^{3N})$  for the antisymmetric (fermionic) subspace of  $L^2(\mathbb{R}^{3N})$ . A particular example of a bosonic wave function is the  $N$ -fold tensor product of some one-particle wave function  $\varphi \in L^2(\mathbb{R}^3)$ ,

$$\begin{aligned}\psi(x_1, x_2, x_3, \dots) &= \varphi^{\otimes N}(x_1, x_2, x_3, \dots) \\ &= \prod_{i=1}^N \varphi(x_i).\end{aligned}\tag{4}$$

A wave function of the form (4) is a perfect *Bose–Einstein condensate*: one says that all bosons have condensed into the orbital  $\varphi$ .

The simplest wave function compatible with fermionic antisymmetry is obtained by picking  $N$  linearly independent one-particle wave function  $\varphi_j \in L^2(\mathbb{R}^3)$ , forming a tensor product and antisymmetrizing this tensor product, resulting in

$$\begin{aligned}\psi(x_1, x_2, x_3, \dots, x_N) &= \frac{1}{N!} \sum_{\sigma \in S_N} \text{sgn}(\sigma) \varphi_{\sigma(1)}(x_1) \varphi_{\sigma(2)}(x_2) \cdots \varphi_{\sigma(N)}(x_N) \\ &= (N!)^{-1/2} \det(\varphi_i(x_j))_{i,j=1,\dots,N}.\end{aligned}\tag{5}$$

A wave function of the form (5) is also called a *Slater determinant*.

## Quantities of Interest

The main physical quantities of interest are the excitation energies, given by the spectrum  $\sigma(H_N)$  of the Hamiltonian, and the time evolution of initial data  $\psi(0) \in L^2(\mathbb{R}^{3N})$  according to the time-dependent Schrödinger equation

$$i\hbar\partial_t\psi(t) = H_N\psi(t), \quad (6)$$

where at every time  $t \in \mathbb{R}$ , we have  $\psi(t) \in L^2(\mathbb{R}^{3N})$ . Since  $H_N$  is invariant under permutation of the particle coordinates  $x_i \in \mathbb{R}^3$ , if  $\psi(0)$  is fermionic (i.e., antisymmetric), so is the solution  $\psi(t)$  at all times, and likewise for the bosonic case. There is a vast body of literature on the dynamics; we refer to Rougerie (2021, Sect. 1.8.2) and Napiórkowski (2021).

If  $E$  is an eigenvalue of the Hamiltonian, i.e., there exists  $\psi_E$  such that

$$H_N\psi_E = E\psi_E,$$

then  $\psi(t) = e^{-iEt/\hbar}\psi_E$  is a stationary solution of the Schrödinger equation, and  $E$  is called an energy level. Coupling to an electromagnetic field may induce transition between such energy levels by absorption or emission of photons. These are among the most common observations in experiments. A particularly important energy level is the *ground state energy*

$$E_N = \inf \sigma(H_N) = \inf_{\psi \in L^2_{s/a}(\mathbb{R}^{3N}), \|\psi\|=1} \langle \psi, H_N \psi \rangle.$$

Under general assumptions on the potential  $V$  and if  $V_{\text{ext}}(\mathbf{x})$  tends to infinity at large  $|\mathbf{x}|$ , the ground state energy  $E_N$  is an eigenvalue of the Hamiltonian (Reed and Simon 2012). All corresponding eigenvectors are called ground states of  $H_N$ . In all of the following, we focus on the computation of the ground state energy, although the concepts are far more generally applicable, e.g., to analyze higher energy levels and the time evolution.

Mathematically, the ground state energy is the best understood quantity, and the further properties are still subject to active investigation. We see the presentation of methods in the following therefore also as a toolbox to enable the analysis of more complicated observable properties (like transport coefficients), or of more complicated physical systems. In the future, they may pave a way to understand phenomena such as superfluidity of Bose–Einstein condensates or justify the widely used Fermi liquid theory for electron systems.

## Article Roadmap

In Subsection “[Scaling Limits](#),” we discuss the most important scaling limits, modelling idealized physical situations. In Subsection “[Second Quantization](#),” we then review second quantization as the basis for the further theory. In section “[Bogoliubov Transformations and Quasi-Free States](#),” we discuss quasi-free states and their role as ground states of quadratic Hamiltonians. In section “[Quasi-free Approximations](#),” we introduce quasi-free approximations, i.e., approximations to interacting many-body systems obtained by restriction of the admitted states to quasi-free states; we will discuss their predictions for the ground state energy. In sections “[Correlation Corrections to the Gross–Pitaevskii Approximation](#)” and “[Correlation Corrections to the Hartree–Fock Approximation](#),” we come to the main topic, correlation corrections (deviations of the ground state from its quasi-free approximation) as perturbations to the quasi-free approximations.

## Scaling Limits

In the following, we focus on two important cases and introduce the scaling limits modelling them. For bosons, a lot of experimental work has been focused on dilute cold atomic gases; the experimental situation can be modelled as the Gross–Pitaevskii scaling limit. For fermions, metals are of particular interest; they can be modelled as being of rather high density, while the interaction has a long range, i.e., mean-field behavior.

Another well-known scaling limit for bosons is the mean-field limit describing high density

systems and corresponding to the choice  $\lambda = N^{-1}$ ,  $\hbar = 1$ . This case is mathematically simpler; we refer to (Boßmann et al. 2021; Seiringer 2011; Spohn 1980) and do not discuss it further. Fermions have also been studied in dilute limits, for which we refer to (Falconi et al. 2021).

### Bosons: Low-Density Scaling Limit (Gross–Pitaevskii Limit)

The model describes  $N$  bosons trapped by the external potential  $V_{\text{ext}}: \mathbb{R}^3 \rightarrow \mathbb{R}$  (e.g., a harmonic oscillator potential  $V_{\text{ext}}(x) = x^2$ ). Acting on  $L_s^2(\mathbb{R}^{3N})$ , its Hamiltonian is

$$H_N = \sum_{i=1}^N (-\Delta_{x_i} + V_{\text{ext}}(x_i)) + N^2 \sum_{i < j}^N V(N(x_i - x_j)). \quad (7)$$

The scaling of the interaction potential with  $N$  realizes a low density (dilute) regime for large  $N$ : particles collide only when their distance is of order  $N^{-1}$  (thinking of  $V$  as compactly supported), whereas the typical distance between them is of order  $N^{-1/3}$ , i.e., much larger. The Planck constant is independent of  $N$  and, by a choice of units,  $\hbar = 1$ . The choice  $\lambda = N^2$  ensures that the expectation values of the kinetic part and the interaction part in a condensate state are of the same order of magnitude. This is the Gross–Pitaevskii scaling limit. By rescaling lengths, (7) is unitarily equivalent (up to an overall factor  $N^2$ ) to a Hamiltonian with unscaled potential for particles on a torus of side length  $N$ . The limit  $N \rightarrow \infty$  then corresponds to a simultaneous large volume and low density limit with density of order  $N^{-2}$ .

An important simplification that retains most of the interesting features of the model is to drop the confining potential  $V_{\text{ext}}$  and instead consider it on the torus  $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$ . It is then convenient to express the model in Fourier (momentum) representation in terms of the creation operators  $a_k^* := a^*(f_k)$  for plane waves  $f_k(x) = e^{ik \cdot x}$ , with momenta  $k \in 2\pi\mathbb{Z}^3$  (See subsection “[Second](#)

[Quantization](#)” for an introduction to the formalism of second quantization.). The number operator can be expressed as  $\mathcal{N} = \sum_{p \in 2\pi\mathbb{Z}^3} a_p^* a_p$ . The Hamiltonian  $H_N$  can be lifted to Fock space by defining

$$H_N = \sum_{p \in 2\pi\mathbb{Z}^3} p^2 a_p^* a_p + \frac{1}{2N} \sum_{p, q, r \in 2\pi\mathbb{Z}^3} \widehat{V}(r/N) a_p^* a_q^* a_{q-r} a_{p+r}. \quad (8)$$

In fact, considering  $L_s^2(\mathbb{T}^{3N})$  as a subspace of  $\mathcal{F}_s$ , we note that  $\mathcal{H}_N \upharpoonright_{L_s^2(\mathbb{T}^{3N})} = H_N$ .

### Fermions: High-Density Scaling Limit (Mean-Field/Semiclassical Limit)

As for boson we drop the external potential and consider  $N$  fermions on the torus  $\mathbb{T}^3$ , which, given the fixed volume of the torus, models a high-density regime. Note that for the noninteracting case  $V = 0$ , the ground state is given by a Slater determinant  $\psi_{B_F}$  of plane waves  $f_k$  with

$$k \in B_F := \{k \in 2\pi\mathbb{Z}^3 \mid |k| \leq k_F := (6\pi^2)^{1/3} N^{1/3}\},$$

which is called the *Fermi ball*. (We assume that the particle number  $N$  is such that all lattice points in the ball are occupied.) One then finds

$$E_N = \langle \psi_{B_F}, H_N \psi_{B_F} \rangle = \hbar^2 \sum_{k \in B_F} |k|^2 = \mathcal{O}(\hbar^2 N^{5/3}).$$

Thinking about the interacting system again, the expectation value of the interaction is of order  $N^2$  (since particles interact in all pairs, the number of which is  $N(N-1)/2$ ). To get kinetic and interaction part of the same magnitude, we choose

$$\lambda := N^{-1} \text{ and the effective Planck constant } \hbar := N^{-1/3}.$$

Note that in the time-dependent Schrödinger Eq. (7),  $\hbar$  also appears together with the time derivative, corresponding to a rescaling of the time variable to semiclassical time scales  $t' = N^{-1/3} t$ ,

the inverse of the typical particle velocity  $k_F = \mathcal{O}(N^{1/3})$ . Expressed in terms of fermionic creation and annihilation operators

$$H_N = \sum_{p \in 2\pi\mathbb{Z}^3} \hbar^2 p^2 a_p^* a_p + \frac{1}{2N} \sum_{p, q, r \in 2\pi\mathbb{Z}^3} \widehat{V}(r) a_p^* a_q^* a_{q-r} a_{p+r},$$

$$\hbar = N^{-1/3}.$$

## Second Quantization

To be self-contained, we review the formalism of *second quantization*, a reformulation of the quantum many-body problem that is convenient for the further discussions. A detailed pedagogical discussion may be found in (Solovej 2014). We introduce bosonic and fermionic *Fock space* by

$$\mathcal{F}_s := \bigoplus_{i=0}^{\infty} L_s^2(\mathbb{T}^{3N}), \quad \mathcal{F}_a := \bigoplus_{i=0}^{\infty} L_a^2(\mathbb{T}^{3N}),$$

where both in the symmetric and antisymmetric case  $L_{s/a}^2(\mathbb{T}^0) := \mathbb{C}$ . Vectors in the Fock spaces can be understood as sequences

$$\psi = (\psi^{(0)}, \psi^{(1)}, \psi^{(2)}, \psi^{(3)}, \dots)$$

where the  $j$ -th element is a  $j$ -particle wave function of the appropriate symmetry/antisymmetry,  $\psi^{(j)} \in L_{s/a}^2(\mathbb{T}^{3j})$ . Fock space becomes a Hilbert space with the inner product  $\langle \psi, \varphi \rangle := \sum_{n=0}^{\infty} \langle \psi^{(n)}, \varphi^{(n)} \rangle$  after completion with respect to the norm induced by this inner product. The operator  $\mathcal{N} : \mathcal{F}_{s/a} \mapsto \mathcal{F}_{s/a}$  defined by

$$\mathcal{N}\psi := (0, \psi^{(1)}, 2\psi^{(2)}, 3\psi^{(3)}, \dots)$$

is called the (*particle*) *number operator*. For any one-particle wave function  $f \in L^2(\mathbb{T}^3)$ , we define the *creation operator* acting in Fock space elementwise by

$$(a^*(f)\psi)^{(n)}(x_1, \dots, x_n) := \sum_{j=1}^n \frac{(\pm 1)^j}{\sqrt{n}} f(x_j) \psi^{(n-1)}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$$

(the negative sign is for the fermionic case, the positive sign for the bosonic case), and the *annihilation operator* (with the same formula for bosons and fermions) by

$$(a(f)\psi)^{(n)}(x_1, \dots, x_n) := \sqrt{n+1} \int_{\mathbb{T}^3} dx \overline{f(x)} \psi^{(n+1)}(x, x_1, \dots, x_n).$$

The creation operator is the adjoint of the corresponding annihilation operator,  $a(f)^* = a^*(f)$  for all  $f \in L^2(\mathbb{T}^3)$ . In the bosonic case, they satisfy the *canonical commutator relations* (CCR)

$$[a(f), a(g)] := a(f)a(g) - a(g)a(f) = 0, \\ [a^*(f), a^*(g)] = 0, [a(f), a^*(g)] = \langle f, g \rangle_{L^2(\mathbb{T}^3)},$$

and in the fermionic case, the *canonical anti-commutator relations* (CAR)

$$\{a(f), a(g)\} := a(f)a(g) + a(g)a(f) = 0, \\ \{a^*(f), a^*(g)\} = 0, \{a(f), a^*(g)\} = \langle f, g \rangle_{L^2(\mathbb{T}^3)}.$$

A particularly important vector in Fock space is the *vacuum*, given by

$$\Omega := (1, 0, 0, 0) \in \mathcal{F}_{s/a}.$$

The vacuum has vanishing particle number,  $\mathcal{N}\Omega = 0$ , and all other states in Fock space can be represented by adding (or “creating”) particles to the vacuum. More precisely, if  $(f_i)_{i \in \mathcal{I}}$  is a basis of the one-particle Hilbert space  $L^2(\mathbb{T}^3)$ , then vectors obtained by applying arbitrarily many creation operators to the vacuum,

$$a^*(f_{i_1}) \dots a^*(f_{i_m}) \Omega,$$

constitute a basis of the respective Fock space  $\mathcal{F}_{s/a}$ . Furthermore, the vacuum is in the kernel of all annihilation operators,  $a(f)\Omega = 0$  for all  $f \in L^2(\mathbb{T}^3)$ .

For  $x \in \mathbb{T}^3$ , one also has the densely defined operator

$$(a_x \psi)^{(n)}(x_1, \dots, x_n) := \sqrt{n+1} \psi^{(n+1)}(x, x_1, \dots, x_n).$$



Formally,  $a_x = a(\delta_x)$ , where  $\delta_x$  is the delta distribution at  $x \in \mathbb{T}^3$ . Instead,  $a_x^*$  can only be defined as an operator on the trivial domain  $\{0\}$ ; we can, however, think of it as an operator-valued distribution satisfying  $a^*(f) = \int f(x) a_x^* dx$ . In normal-ordered products of creation and annihilation operators (i.e., products in which all creation operators stand to the left of the annihilation operators), we may also think of it as defined in quadratic form by  $\langle \varphi, a_x^* \psi \rangle := \langle a_x \varphi, \psi \rangle$  for  $\varphi, \psi \in \mathcal{F}$ .

## Bogoliubov Transformations and Quasi-free States

In this section, we introduce quasi-free state following closely the presentation of (Solovej 2014).

We denote the complex conjugation map by  $J : L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3)$ , and define the *generalized creation and annihilation operators*, for both the fermionic and the bosonic case, by

$$A(f \oplus Jg) := a(f) + a^*(g).$$

Let  $F \in L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$ . The property

$$A^*(F) = A(JF),$$

where the block operator

$$J = \begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix}$$

acts on

$$L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3),$$

represents the fact that creation operators are the adjoint of the annihilation operators. Furthermore, for the bosonic case, the CCR are equivalent to the commutator conditions

$$[A(f_1), A^*(f_2)] = \langle f_1, Sf_2 \rangle_{L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)},$$

where the block operator

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

acts on

$$L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3).$$

For the fermionic case, the CAR are equivalent to the anticommutator conditions

$$\{A(F_1), A^*(F_2)\} = \langle F_1, F_2 \rangle_{L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)}.$$

We may think of the operators  $a^*(f)$  as enabling us to describe a physical system in terms of the creation of particles. A Bogoliubov map (or *Bogoliubov transformation*) will enable us to represent a system in terms of the creation of “quasiparticles,” for which the corresponding operators are linear combinations of creation and annihilation operators such that the (anti)commutator relations for the quasiparticles have exactly the same form as those of the original particles. This is formalized by the following definition:

### Definition 1 (Bogoliubov map)

A linear bounded isomorphism  $\mathcal{V} : L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$  is called a *bosonic Bogoliubov map* if for all  $F, F_1, F_2 \in L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$  we have

$$[A(\mathcal{V}F_1), A^*(\mathcal{V}F_2)] = \langle F_1, SF_2 \rangle_{L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)}$$

$$\text{and } A^*(\mathcal{V}F) = A(\mathcal{V}JF).$$

Likewise, it is called a *fermionic Bogoliubov map* if

$$\{A(\mathcal{V}F_1), A^*(\mathcal{V}F_2)\} = \langle F_1, F_2 \rangle_{L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)}$$

$$\text{and } A^*(\mathcal{V}F) = A(\mathcal{V}JF).$$

One says that a Bogoliubov map  $\mathcal{V}$  has a *unitary implementation* on the respective Fock space if there exists a unitary  $U_{\mathcal{V}} : \mathcal{F}_{s/a} \rightarrow \mathcal{F}_{s/a}$  such that

$$U_{\mathcal{V}} A(F) U_{\mathcal{V}}^* = A(\mathcal{V}F). \quad (9)$$

One can prove (Shale–Stinespring condition, (Solovej 2014, Theorem 9.5) that a Bogoliubov map is unitarily implementable if and only if in its block decomposition

$$\mathbf{v} = \begin{pmatrix} U & JVJ \\ V & JUJ \end{pmatrix},$$

the block  $V$  is a Hilbert–Schmidt operator. The reader may easily convince herself that this corresponds to the number of particles in the transformed vacuum being finite,  $\langle U_{\mathbf{v}}\Omega, \mathcal{N}U_{\mathbf{v}}\Omega \rangle < \infty$ . One may also verify that the transformed vacuum  $U_{\mathbf{v}}\Omega$  is indeed the vacuum for the annihilation operators of the quasiparticles,  $a_{\text{quasi}}(f) := A(\mathbf{v}f \oplus 0)$ . The quasiparticle vacua are also called quasi-free states.

### Definition 2 (Quasi-free state)

Any state  $\psi \in \mathcal{F}_{\text{s/a}}$  that can be obtained from the vacuum  $\Omega = (1, 0, 0, \dots) \in \mathcal{F}_{\text{s/a}}$  by application of a unitary map that implements a Bogoliubov transformation is called a *quasi-free state*.

An important example for a unitarily implemented Bogoliubov map (or Bogoliubov transformation) is provided by operators on  $\mathcal{F}_{\text{s/a}}$  of the form

$$T(\eta) := \exp \left( \sum_{i,j \in \mathcal{I}} \eta_{i,j} a^*(f_i) a^*(f_j) - \text{h.c.} \right), \quad (10)$$

where  $\eta_{i,j} := \langle f_i, \eta f_j \rangle$  are the matrix elements of a Hilbert–Schmidt operator  $\eta : L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3)$ . In particular, states of the form  $T(\eta)\Omega$  are examples of quasi-free states.

The following theorem is very useful in the computation of expectation values and explains the importance of quasi-free states.

### Theorem 1 (Wick theorem)

Let  $\psi \in \mathcal{F}_{\text{s/a}}$  be a quasi-free state and  $F_1, \dots, F_{2m} \in L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$ . Then, for an even number of operators, with the positive sign for bosons and the alternating sign for fermions,

$$\begin{aligned} & \langle \psi, A(F_1) \cdots A(F_{2m}) \psi \rangle \\ &= \sum_{\sigma \in P_{2m}} (\pm 1)^{\sigma} \langle \psi, A(F_{\sigma(1)}) A(F_{\sigma(2)}) \psi \rangle \\ & \quad \cdots \langle \psi, A(F_{\sigma(2m-1)}) A(F_{\sigma(2m)}) \psi \rangle \end{aligned}$$

and for an odd number of operators,

$$\langle \psi, A(F_1) \cdots A(F_{2m-1}) \psi \rangle = 0,$$

where  $P_{2m}$  is the set of pairings (a subset of the set of permutations),

$$\begin{aligned} P_{2m} = \{ \sigma \in S_{2m} \mid & \sigma(2j-1) < \sigma(2j+1) \text{ for } j = 1, \dots, m-1 \\ & \text{and } \sigma(2j-1) < \sigma(2j) \text{ for } j = 1, \dots, m \}. \end{aligned}$$

One can furthermore show that quasi-free states are the only Fock space vectors satisfying the Wick theorem, so that validity of the Wick theorem is sometimes also taken as the definition of quasi-free states. Note also that a Fock space vector (up to multiplication by a phase) is completely determined by all its expectation values of products of creation and annihilation operators; therefore a quasi-free state is completely determined by its *one-particle reduced density matrix*  $\gamma^{(1)}(x, y) = \langle \psi, a_y^* a_x \psi \rangle$  and its *pairing density*  $\alpha(x, y) := \langle \psi, a_y a_x \psi \rangle$ .

**Fermionic Case:** In the case of fermions, the pairing density  $\alpha$  is identified with the wave function of Cooper pairs responsible for superconductivity.

Fermionic quasi-free states with vanishing pairing density  $\alpha = 0$  can be shown to be Slater determinants as follows: if  $\psi \in \mathcal{F}_{\text{a}}$  is quasi-free, we have the relation  $\gamma^2 + \alpha J \alpha J = \gamma$  (Solovej 2014, Eq. 77), which for  $\alpha = 0$  implies that  $\gamma$  is a projection; by the singular value decomposition, one finds  $\gamma = \sum_{j=1}^m |f_j\rangle \langle f_j|$ , where  $|f_j\rangle \langle f_j|$  is the orthogonal projection on  $f_j \in L^2(\mathbb{T}^3)$ ; we may then construct the Slater determinant  $a^*(f_1) \cdots a^*(f_m) \Omega \in \mathcal{F}_{\text{a}}$  which by the uniqueness of a Fock space vector given its expectation values, equals  $\psi$  up to multiplication by a phase.

**Bosonic Case:** In the bosonic case, it is natural to include a further degree of freedom in the class of states for which expectation values can be easily computed. For this purpose, given any  $\varphi \in L^2(\mathbb{T}^3)$ , we define the *Weyl operator*

$$W(\varphi) := \exp(a^*(\varphi) - a(\varphi)). \quad (11)$$

One has the relations

$$W^*(\varphi)a(g)W(\varphi) = a(g) + \langle g, \varphi \rangle \quad (12)$$

and

$$a(g)W(\varphi)\Omega = \langle g, \varphi \rangle W(\varphi)\Omega;$$

the latter shows that *coherent states*  $W(\varphi)\Omega$  are eigenvectors of all annihilation operators. Coherent states can be seen as an idealized description of a Bose–Einstein condensate; in fact  $\gamma(x; y) = \varphi(x)\varphi(y)$ .

The natural class of states to consider for bosonic systems is therefore the class of *quasi-free states with condensate*, i.e., all vectors in  $\mathcal{F}_s$  of the form

$$W(\varphi)U_{\mathcal{V}}\Omega.$$

Expectation values of quasi-free states with condensate may be evaluated in terms of  $\varphi$ ,  $\gamma$ , and  $\alpha$ , by first employing the identity (12) and then the Wick theorem.

A comparison of the typical states for fermionic and bosonic systems is given in Table 1.

### Quadratic Hamiltonians

Bogoliubov maps can be used to compute the spectrum of quadratic Hamiltonians. In the following, we sketch the solution theory (or *diagonalization*) of quadratic Hamiltonians, following (Nam et al., 2016a).

Let  $\mathcal{A} : L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3) \rightarrow L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$  be a self-adjoint trace-class operator. Then, with  $(F_i)_{i \in \mathcal{I}}$ , an orthonormal basis of  $L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)$ , the operator

$$H_{\mathcal{A}} := \sum_{i,j \in \mathcal{I}} \langle F_i, \mathcal{A} F_j \rangle_{L^2(\mathbb{T}^3) \oplus L^2(\mathbb{T}^3)} A^*(F_i) A(F_j) : \mathcal{F}_{s/a} \rightarrow \mathcal{F}_{s/a}, \quad (13)$$

is called a *quadratic Hamiltonian*, or also the *second quantization of  $\mathcal{A}$* . If  $U_{\mathcal{V}}$  is the unitary implementation of a Bogoliubov map  $\mathcal{V}$ , then

$$U_{\mathcal{V}} H_{\mathcal{A}} U_{\mathcal{V}}^* = H_{\mathcal{V} \mathcal{A} \mathcal{V}^*}.$$

Under basic assumptions on  $\mathcal{A}$ , one can find a Bogoliubov map  $\mathcal{V}$  such that

$$\mathcal{V} \mathcal{A} \mathcal{V}^* = \begin{pmatrix} \xi & 0 \\ 0 & J \xi J \end{pmatrix}, \quad (14)$$

where  $\xi : L^2(\mathbb{T}^3) \mapsto L^2(\mathbb{T}^3)$  is diagonal. (For the setting in which  $L^2(\mathbb{T}^3)$  is replaced by a finite-dimensional space, the existence in the bosonic case follows from Williamson (1936), assuring that the matrix  $\mathcal{A}$  can be diagonalized by conjugation with a symplectic matrix, the symplecticity needed due to the presence of  $\mathcal{S}$  in the CCR.) This means that for some orthonormal basis  $(f_m)_m$  of  $L^2(\mathbb{T}^3)$ , we have

$$U_{\mathcal{V}} H_{\mathcal{A}} U_{\mathcal{V}}^* = \sum_{m \geq 0} \langle f_m, \xi f_m \rangle_{L^2(\mathbb{T}^3)} a^*(f_m) a(f_m) + \frac{1}{2} \operatorname{tr} \xi. \quad (15)$$

Since the spectrum of  $a^*(f_m) a(f_m)$  is given by the natural numbers  $\mathbb{N}_+ := \mathbb{N}$  (recall the algebraic solution of the harmonic oscillator) in the bosonic case and  $\mathbb{N}_- := \{0, 1\}$  in the fermionic case, the spectrum can then be read off to be

**Correlation Corrections as a Perturbation to the Quasi-free Approximation in Many-Body Quantum Systems, Table 1** Comparison of bosonic and fermionic

theory (See Sections 4 and 5 for the discussion of non-trivial correlation corrections)

|                         | Bosons on $\mathcal{F}_s$                                                                         | Fermions on $\mathcal{F}_a$                                                                                |
|-------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Uncorrelated state      | Coherent state $W(N^{1/2})\Omega$                                                                 | Does not exist because of antisymmetry requirement                                                         |
| Quasi-free state        | $W(N^{1/2})T\Omega$ , where $T \sim \exp.(a^* a^* - \text{h. c.})$ is a Bogoliubov transformation | Slater determinant $R\Omega$ , with $R$ a particle-hole transformation                                     |
| Nontrivial correlations | $W(N^{1/2})T e^A \Omega$ with $A \sim (a^* a^* a + \text{h.c.})$                                  | $RT\Omega$ with $T \sim \exp(b^* b^* + \text{h.c.})$ , with approximate emergent bosons $b^* \sim a^* a^*$ |

$$\begin{aligned} \sigma(H_{\mathcal{A}}) &= \sum_{m \geq 0} \sum_{n_m \in \mathbb{N}_{\pm}} \langle f_m, \xi f_m \rangle_{L^2(\mathbb{T}^3)} n_m \\ &\quad + \frac{1}{2} \operatorname{tr} \xi. \end{aligned} \quad (16)$$

In this sense, quadratic Hamiltonians are diagonalizable by Bogoliubov transformations. In particular, the ground states of quadratic Hamiltonians are quasi-free.

At this point, we have all the tools at hand that are needed to define the quasi-free approximation of a many-body system, in the following discussed for the example of a dilute bosonic system (Gross–Pitaevskii approximation) and a high-density fermionic system (Hartree–Fock approximation).

## Quasi-free Approximations

In the simplest case, the quasi-free approximation is obtained by considering the ground state energy variational problem  $E_N = \inf_{\psi} \langle \psi, H_N \psi \rangle$  and restricting the set of permitted  $\psi$  to quasi-free states (in the bosonic case also with condensate). Of course, the ground state energy  $E_N$  is always smaller than its quasi-free approximation; the dif-

ference is called *correlation energy*. However, the quasi-free approximation permits the explicit computation of the expectation value using the Wick theorem, leading to a functional depending on  $\gamma = \gamma^{(1)}$ ,  $\alpha$ , and (for bosons)  $\varphi$ . The idea of reduction to quasi-free states can be generalized to the dynamics, for which we refer to (Bach et al. 2016; Benedikter et al. 2018; Lubich 2008).

We now discuss the bosonic and fermionic case separately in more detail.

### Bosons: Gross–Pitaevskii Approximation

Consider the bosonic quasi-free state  $T(\eta)W(N^{1/2}\varphi)\Omega$ , with  $T(\eta)$  as in (10) and  $W(N^{1/2}\varphi)$  as in (11). Define the integral kernel of the operator  $\eta$  in the Bogoliubov transformation  $T(\eta)$  to be of the form  $\eta(x; y) = -N(1 - f)(N(x - y))\varphi(x)\varphi(y)$ , where  $f$  is the solution of the scattering Eq. (1), and take  $\varphi \in L^2(\mathbb{R}^3)$ . The energy of  $H_N$  (see (7)) in the state  $T(\eta)W(N^{1/2}\varphi)\Omega$  is

$$\begin{aligned} &\langle T(\eta)W(N^{1/2}\varphi)\Omega, H_N T(\eta)W(N^{1/2}\varphi)\Omega \rangle \\ &= N\mathcal{E}_{\text{GP}}(\varphi) + \mathcal{O}(1) \end{aligned} \quad (17)$$

where  $\mathcal{E}_{\text{GP}}$  is the Gross–Pitaevskii functional

$$\mathcal{E}_{\text{GP}}(\varphi) := \int dx \left[ |\nabla \varphi(x)|^2 + V_{\text{ext}}(x)|\varphi(x)|^2 + 4\pi a |\varphi(x)|^4 \right]. \quad (18)$$

If we call  $\varphi_0$  the minimizer of  $\mathcal{E}_{\text{GP}}$  over all normalized  $\varphi \in L^2(\mathbb{R}^3)$ , the quasi-free state  $T(\eta)W(N^{1/2}\varphi_0)\Omega$  gives the smallest energy, given the choice of  $\eta$  in terms of  $\varphi_0$  as above. The transformation  $T(\eta)$  modifies the coherent state on a length scale of order  $N^{-1}$  in terms of the solution of the two-body scattering equation. We refer to Benedikter et al. (2021a, Appendix A) for an explicit computation of (17). Indeed the quasi-free state  $TW(N^{1/2}\varphi_0)$  captures the leading order of the ground state energy: it has been proved in Boccato et al. (2018, 2020); Brennecke et al. (2022); Lieb et al. (2000); Nam et al. (2016b, 2020) that

$$E_N = N \min_{\substack{\varphi \in L^2(\mathbb{R}^3) \\ \|\varphi\|_2=1}} \mathcal{E}_{\text{GP}}(\varphi) + \mathcal{O}(1). \quad (19)$$

The Gross–Pitaevskii approximation provides a one-body effective theory for the ground state energy of the Bose gas (and for its dynamics). Notice that while the original many-body theory is linear, the Gross–Pitaevskii approximation is nonlinear. Observe also that it shows a universal behavior, because the scattering length  $a$  is the only parameter associated to  $V$  appearing in the ground state energy (details of  $V$  do not matter). If we consider the system confined in the torus and

described by (8), the leading order of the ground state energy reduces to

$$E_N = 4\pi aN + \mathcal{O}(1) \quad (20)$$

since the minimizer of  $\mathcal{E}_{\text{GP}}$  is  $\varphi_0 = 1$ .

While quasi-free states resolve the leading order of the ground state energy, they cannot capture the second order (Erdos et al. 2008; Napiórkowski et al. 2018a; b). In particular, in (Erdos et al. 2008), an approximate minimization over the choice of  $\eta$  is performed; the corresponding trial state, however, does not yield the expected value of the ground state

energy at second order (the correct second order is recovered only in regimes of weak interactions).

### Fermions: Hartree–Fock Approximation

The quasi-free approximation for fermions is given by the variational problem

$$E_N^{\text{qf}} := \inf_{\psi \text{ is a Bogoliubov map}} \langle U\psi\Omega, H_N U\psi\Omega \rangle.$$

Computing the expectation value by the Wick theorem, one finds (Hainzl and Seiringer 2016) the functional, expressed in terms of  $\gamma$  and  $\alpha$  associated to the vector  $U\psi\Omega$ ,

$$\begin{aligned} \mathcal{E}(\gamma, \alpha) = & \text{tr}(-\hbar^2 \Delta \gamma) + \frac{1}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} dx dy V(x-y) \left( \gamma(x; x) \gamma(y; y) - |\gamma(x; y)|^2 \right) \\ & + \frac{1}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} dx dy V(x-y) |\alpha(x, y)|^2. \end{aligned}$$

If the Fourier transform of  $V$  is nonnegative,  $\widehat{V}(k) \geq 0$  for all  $k \in 2\pi\mathbb{Z}^3$ , it is preferable to have

$\alpha = 0$  in the minimizer. In that case, the functional simplifies to the *Hartree–Fock functional*

$$\mathcal{E}^{\text{HF}}(\gamma) = \text{tr}(-\hbar^2 \Delta \gamma) + \frac{1}{2} \int_{\mathbb{T}^3 \times \mathbb{T}^3} dx dy V(x-y) \left( \gamma(x; x) \gamma(y; y) - |\gamma(x; y)|^2 \right).$$

This is the functional we consider as the starting point for the discussion of correlation corrections. According to (Graf and Solovej 1994) (there proven for the thermodynamic limit, but easily found to be applicable also for the mean-field/semiclassical scaling limit), the Hartree–Fock energy  $E^{\text{HF}} := \inf_{\gamma^2=\gamma} \mathcal{E}^{\text{HF}}(\gamma)$  approximates the ground state energy,

$$E_N = E_N^{\text{HF}} + o(1) \text{ as } N \rightarrow \infty.$$

It can be shown that in the mean-field/semiclassical scaling limit, the Hartree–Fock minimizer is, like in the noninteracting case, given by the plane waves in the Fermi ball (Benedikter et al. 2014, Appendix A). This is a special property of the translation invariant setting, but even in the translation invariant case not true for other

scalings (Gontier and Lewin 2019) (although the energy deviation may be exponentially small (Gontier et al. 2019)).

In the following, we will continue to focus on the mean-field/semiclassical scaling limit with  $\widehat{V} \geq 0$  (so  $\alpha = 0$ ), so that the Slater determinant of plane waves  $\psi_{B_F}$  will be the starting point for the addition of quantum correlations.

### Correlation Corrections to the Gross–Pitaevskii Approximation

We consider the system on the torus  $\mathbb{T}^3$ . We discuss how to modify the quasi-free theory to obtain the second order to (20) and the low-energy eigenvalues of (8). This is the content of the theorem below, proved in Boccato et al. (2019).

**Theorem 2 (Correlation Energy and Excitation Spectrum)**

Let  $V \in L^3(\mathbb{R}^3)$  be nonnegative, spherically symmetric, and compactly supported. Then,

for  $N \rightarrow \infty$ , the ground state energy is given by

$$E_N = 4\pi(N-1)a + e_\Lambda a^2 - \frac{1}{2} \sum_{p \in 2\pi\mathbb{Z}^3 \setminus \{0\}} \left[ p^2 + 8\pi a - \sqrt{|p|^4 + 16\pi a p^2} - \frac{(8\pi a)^2}{2p^2} \right] + \mathcal{O}(N^{-1/4}). \quad (21)$$

Here we defined

$$e_\Lambda := 2 - \lim_{M \rightarrow \infty} \sum_{\substack{p \in \mathbb{Z}^3 \setminus \{0\}: \\ |p_1|, |p_2|, |p_3| \leq M}} \frac{\cos(|p|)}{p^2}.$$

Moreover, the spectrum of  $H_N - E_N$  below a threshold  $\zeta$  consists of eigenvalues given, in the limit  $N \rightarrow \infty$ , by

$$\sum_{p \in 2\pi\mathbb{Z}^3 \setminus \{0\}} n_p \sqrt{|p|^4 + 16\pi a p^2} + \mathcal{O}(N^{-1/4}(1 + \zeta^3)). \quad (22)$$

Here  $n_p \in \mathbb{N}$  for all  $p \in 2\pi\mathbb{Z}^3 \setminus \{0\}$  and  $n_p \neq 0$  for finitely many  $p \in 2\pi\mathbb{Z}^3 \setminus \{0\}$  only.

The summand in the second line of (21) behaves as  $p^{-4}$  for large  $p$ , so that the sum is finite. The term  $e_\Lambda a^2$  originates as a boundary contribution. Like the second order of the ground state energy, the spacing between eigenvalues is a quantity of order one. The excitation spectrum (22) is given by a sum of approximately noninteracting harmonic oscillators (the second term on the right-hand side in (16) corresponds to the ground state energy (21)). For each  $p$ ,  $\mathcal{E}(p) = \sqrt{|p|^4 + 16\pi a p^2}$  represents the dispersion relation (energy) of excitations. For small momenta,  $\mathcal{E}(p) = \sqrt{16\pi a}|p|(1 + \mathcal{O}(p^2))$  is linear. This has important physical consequences: the linear behavior (completely different from the quadratic dispersion of free particles) is a signature of superfluidity, according to Landau's criterion

(Landau 1941). Notice again that the scattering length is the only parameter related to the interaction potential which appears in the ground state energy and in the excitation spectrum.

The strategy to prove Theorem 2 is inspired by Bogoliubov theory. We construct a quadratic effective Hamiltonian approximating the quartic operator (8) and obtain (21) and (22) by diagonalization, with the method outlined in section “Quadratic Hamiltonians.” The crucial point is that we need to correct Bogoliubov's quadratic Hamiltonian and go beyond the quasi-free approximation. To explain this point, it is instructive to review Bogoliubov theory.

Developed in Bogolubov (1947), *Bogoliubov theory* originally aimed at explaining microscopically the phenomenon of superfluidity in liquid Helium. By heuristic arguments and physical considerations, Bogoliubov proposes a quadratic effective Hamiltonian in the form (13) to be diagonalized through a Bogoliubov transformation as in (15). The first observation in Bogolubov (1947) is that if Bose-Einstein condensation occurs, the number of particles with momentum  $p = 0$  is order  $N$ , i.e.,  $a_0^* a_0 \simeq N$ . This justifies the replacement of  $a_0$  and  $a_0^*$  in the Hamiltonian by  $N^{1/2}$ . Such formal substitution corresponds to conjugating the Hamiltonian with the Weyl operator  $W(N^{1/2}) = \exp(a^*(N^{1/2}) - a(N^{1/2}))$ , as defined in (11). With the substitution (or the shift (12)),  $H_N$  decomposes in a sum of contributions that are quadratic, cubic, and quartic in creation and annihilation operators, plus a contribution that does not contain creation and annihilation operators.



Cubic and quartic contributions present a pre-factor  $N^{-1/2}$  and  $N^{-1}$  respectively, so they are formally of lower order. Bogoliubov neglects them, remaining with a Hamiltonian of the form

$$H_{\text{Bog}} \simeq \frac{N}{2} \widehat{V}(0) + \sum_{p \in 2\pi\mathbb{Z}^3 \setminus \{0\}} (p^2 + \widehat{V}(p/N)) a_p^* a_p + \sum_{p \in \mathbb{Z}^3 \setminus \{0\}} \frac{\widehat{V}(p/N)}{2} (a_p^* a_{-p}^* + a_p a_{-p}). \quad (23)$$

Being quadratic,  $H_{\text{Bog}}$  can be explicitly diagonalized through a Bogoliubov transformation  $T(\kappa)$  of the form (10) to obtain the ground state energy and the excitation spectrum. The explicit form of the operator  $\kappa$  in  $T(\kappa)$  can be obtained by following the method outlined in section “[Quadratic Hamiltonians](#).” (One finds a  $\kappa$  different from the  $\eta$  used in (17).). However, the ground state energy and the excitation spectrum of (23) do not coincide with (21) and (22); they have the same form, but instead of the scattering length, only its first Born approximation appears (The Born series of the scattering length is obtained substituting recursively the solution of (1) into

(2); in terms of the Fourier transformed  $\widehat{V}$ , it reads

$$8\pi a = \int V(x) dx - \frac{1}{2} \int dp \frac{\widehat{V}^2(p)}{p^2} + \frac{1}{2^2} \int dp \widehat{V}(p) \frac{1}{p^2} \int dr \widehat{V}(r-p) \times \frac{\widehat{V}(r)}{r^2} + \dots \quad (24)$$

Bogoliubov comments that the correct results should contain the scattering length and performs an ad hoc substitution. The contributions missing to obtain the scattering length in the ground state energy and the spectrum are therefore to be extracted from the cubic and quartic terms that Bogoliubov neglected. (Note that the ground state energy of the Bogoliubov Hamiltonian does not coincide with  $E_N$  also at leading order. This is due to the fact that quartic terms (which contribute at order  $N$ ) are taken into account in (17) but are neglected in Bogoliubov’s approximation before the diagonalization.)

The key idea in the proof of Theorem 2 is to construct a unitary operator  $e^A$ , given by the exponential of a *cubic expression in creation and annihilation operators* as

$$A = \frac{1}{\sqrt{N}} \sum_{r \in P_H, v \in P_L} \widehat{\eta}_r [\sinh(\widehat{\eta}_v) a_{r+v}^* a_{-r}^* a_{-v}^* + \cosh(\widehat{\eta}_v) a_{r+v}^* a_{-r}^* a_v - \text{h.c.}] \quad (25)$$

where  $P_H$  and  $P_L$  are suitable intervals (high and low) of momenta. Conjugation with  $e^A$  extracts the main (i.e., order 1) contributions from cubic and quartic terms and produces new quadratic and constant terms. This leads to a new quadratic Hamiltonian, renormalized in the sense that every instance of the potential is corrected into the scattering length. This Hamiltonian can be diagonalized (through a further Bogoliubov transformation  $T(\tau)$  as in (10)) and yields the correct ground state energy and the excitation spectrum.

The ground state vector can be approximated (up to a phase) by a state of the form  $W(N_0^{1/2}) T(\eta) e^A T(\tau) \Omega$ , with  $N_0$  the number of particles with momentum  $p = 0$ . Due to the presence of  $e^A$ , the state is not quasi-free.

While Bogoliubov transformations act linearly on creation and annihilation operators (see the explicit relations (12) and (9)) and in closed form, the cubic unitary does not. To compute its action on  $H_N$ , it is necessary to expand in a series

of commutators, identify the main contributions, and control the remaining terms in the series. (We omitted here many other aspects of the proof and refer to (Boccatto 2021; Cenatiempo 2019) for more detailed reviews.)

## Correlation Corrections to the Hartree–Fock Approximation

Correlations corrections to Hartree–Fock theory can be described through collective bosonization of particle-hole pair excitations. This idea dates back to the (non-collective) bosonization approach of (Sawada 1957; Sawada et al. 1957), which is one of at least three different formulations of the *random phase approximation*.

The problem of computing correlation corrections to the Hartree–Fock energy of the electron gas – a system of fermions interacting through the repulsive Coulomb potential, modelling electrons in metals – has been one of the founding problems of theoretical condensed matter physics. In fact, it was noticed early on (Heisenberg 1947; Wigner 1934) that the standard quantum-mechanical perturbation theory produced corrections to the Hartree–Fock ground state energy that diverge logarithmically in the low-energy region (infrared divergence) already at second order, and become more severely divergent with every order of the expansion. Macke (1950) understood that this is a problem only of the

perturbative expansion, and proposed that a partial resummation of the perturbative series should be better behaved. Bohm and Pines (Bohm and Pines 1953) then proposed a physical picture for the behavior of the electron gas: the long range of the Coulomb interaction leads to the formation of collective oscillations of all electrons; these oscillations screen the potential so that only a weak, short-ranged interaction between the electrons remains. The collective oscillations have a frequency that is observable as the plasma frequency. Gell–Mann and Brueckner showed that this picture is in agreement with the analysis of Macke, and pushed the resummation further to obtain also the next smaller contribution to the ground state energy. Finally, Sawada, Brueckner, Fukuda, and Brout in an already mentioned paper (Sawada et al. 1957) provided an interpretation in terms of the bosonic behavior of particle-hole pairs. Rigorously, the state of the art is the following theorem:

### Theorem 3 (Correlation Energy)

Let  $V \in L^1(\mathbb{T}^3)$  with  $\widehat{V} \geq 0$  and  $\widehat{V}(k)|k| < \infty$ . For  $k_F > 0$ , let  $N = |B_F| = |\{k \in \mathbb{Z}^3 : |k| \leq k_F\}|$ . Then there exists  $\alpha > 0$  such that, for  $k_F \rightarrow \infty$ ,

$$E_N = E_N^{\text{HF}} + E_N^{\text{RPA}} + \mathcal{O}\left(N^{-1/3-\alpha}\right) \quad (26)$$

where, with  $\kappa = (6\pi^2)^{1/3}$ ,

$$E_N^{\text{RPA}} = \hbar\kappa \sum_{k \in 2\pi\mathbb{Z}^3} |k| \left( \frac{1}{\pi} \int_0^\infty \log \left[ 1 + 2\pi\kappa \widehat{V}(k) (1 - \lambda \arctan(1/\lambda)) \right] d\lambda - \frac{\pi}{2} \kappa \widehat{V}(k) \right).$$

This theorem has been proven in the series of papers (Benedikter et al. 2020) (upper bound on  $E_N$ ), (Benedikter et al. 2021a) (lower bound for small potential), and (Benedikter et al. 2021b) (lower bound for general potential) using a collective bosonization method. A similar result has been also obtained by a non-collective method in (Christiansen et al. 2021). To second order in  $V$ , the RPA formula has been obtained earlier in

(Hainzl et al. 2020) via conventional perturbation theory.

## Theory of Correlation Corrections for Fermions

Theorem 3 is proven by adding a description of quantum correlations to the quasi-free minimizer, i.e., to the Slater determinant of plane waves. The Bogoliubov transformation describing the quasi-

free state is of a special form, namely a *particle-hole transformation*  $R : \mathcal{F}_a \rightarrow \mathcal{F}_a$ . To define  $R$ , it suffices to specify its action on the vacuum and the creation operators:

$$R^* a_p^* R := \begin{cases} a_p^* & \text{for } p \in B_F^c \\ a_p & \text{for } p \in B_F \end{cases}, \quad R\Omega := \prod_{k \in B_F} a_k^* \Omega.$$

We transform the Hamiltonian to describe only the effect of nontrivial correlations by setting

$$\begin{aligned} \mathcal{H}_{\text{corr}} &:= R^* \mathcal{H}_N R - E_N^{\text{HF}} \\ &= \mathbb{H}_0 + Q_B + \mathcal{E}_1 + \mathcal{E}_2 + \mathbb{X}. \end{aligned} \quad (27)$$

The correlation Hamiltonian is written using the *particle-hole pair creation operator*

$$b^*(k) := \sum_{p \in B_F^c \cap (B_F + k)} a_p^* a_{p-k} \quad (28)$$

and  $\mathfrak{D}(k)^* := \sum_{p \in B_F^c \cap (B_F^c + k)} a_p^* a_{p-k} - \sum_{h \in B_F \cap (B_F - k)} a_h^* a_{h+k}$  in terms of

$$\begin{aligned} \mathbb{H}_0 &:= \sum_{k \in 2\pi\mathbb{Z}^3} e(k) a_k^* a_k, \quad \text{with } e(k) := |\hbar^2 |k|^2 - \kappa^2|, \\ Q_B &:= \frac{1}{N} \sum_{k \in \Gamma^{\text{nor}}} \widehat{V}(k) [b^*(k) b(k) + b^*(-k) b(-k) + b^*(k) b^*(-k) + b(-k) b(k)], \\ \mathcal{E}_1 &:= \frac{1}{2N} \sum_{k \in \Gamma^{\text{nor}}} \widehat{V}(k) [\mathfrak{D}(k)^* \mathfrak{D}(k) + \mathfrak{D}(-k)^* \mathfrak{D}(-k)], \\ \mathcal{E}_2 &:= \frac{1}{N} \sum_{k \in \Gamma^{\text{nor}}} \widehat{V}(k) [\mathfrak{D}(k)^* b(k) + \mathfrak{D}(-k)^* b(-k) + \text{h.c.}], \\ \mathbb{X} &:= -\frac{1}{2N} \sum_{k \in 2\pi\mathbb{Z}^3} \widehat{V}(k) \left[ \sum_{p \in B_F^c \cap (B_F + k)} a_p^* a_p + \sum_{h \in B_F \cap (B_F^c - k)} a_h^* a_h \right]. \end{aligned}$$

Note that we have introduced the set  $\Gamma^{\text{nor}}$  of all momenta  $k = (k_1, k_2, k_3) \in 2\pi\mathbb{Z}^3$  satisfying  $k_3 > 0$  or  $(k_3 = 0 \text{ and } k_2 = 0)$  or  $(k_2 = k_3 = 0 \text{ and } k_1 > 0)$ , chosen such that  $\Gamma^{\text{nor}} \cap (-\Gamma^{\text{nor}}) = \emptyset$  and  $\Gamma^{\text{nor}} \cup (-\Gamma^{\text{nor}}) = 2\pi\mathbb{Z}^3 \setminus \{0\}$ .

The summands  $\mathbb{H}_0$  and  $\mathbb{X}$ , being quadratic in fermionic creation and annihilation operators, constitute the effective Hamiltonian of Hartree–Fock theory. For  $\mathbb{H}_0 + \mathbb{X}$ , the ground state is given by the vacuum vector  $\Omega$  (corresponding to the Slater determinant  $R\Omega$  before the particle-hole transformation). We may now think of  $Q_B$ ,  $\mathcal{E}_1$ , and  $\mathcal{E}_2$  as perturbations of the Hartree–Fock approximation. Writing a general state  $\psi \in L_a^2(\mathbb{T}^{3N}) \subset \mathcal{F}_a$  as  $\psi = R\xi$ , we can think of  $\xi \in \mathcal{F}_a$  as describing the nontrivial quantum correlations of the state, potentially lowering the energy compared to the Slater determinant as a trial state.

The description of correlations that leads to the random phase approximation is based on the observation that  $\mathbb{X}$ ,  $\mathcal{E}_1$ , and  $\mathcal{E}_2$  have contributions to the energy that can be estimated to be small (Benedikter et al. 2021b), whereas the larger terms  $\mathbb{H}_0 + Q_B$  have an approximate description as a *bosonic* quasi-free theory. In fact, note that the operators  $b(k)$  and  $b^*(k)$  satisfy approximate bosonic CCR:

$$\begin{aligned} [b(k), b(l)] &= 0 = [b^*(k), b^*(l)], \\ [b(k), b^*(l)] &= \delta_{k,l} n_k^2 + \mathcal{E}(k, l), \end{aligned}$$

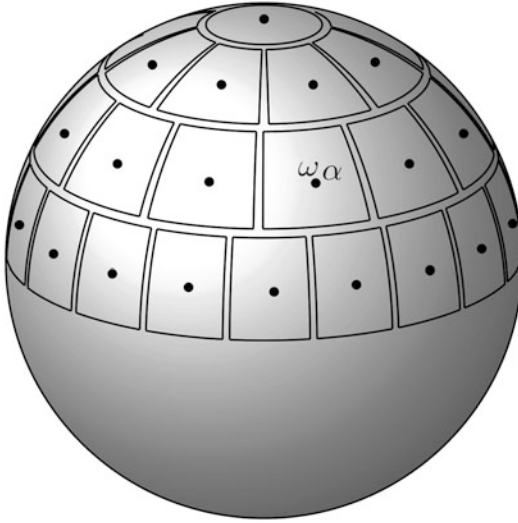
where  $n_k^2$  is a normalization constant and the operator  $\mathcal{E}(k, l)$  may be shown to be a small error in states close to the ground state. Note furthermore that  $Q_B$  is quadratic in the approximately bosonic operators. The missing ingredient is the description of  $\mathbb{H}_0$  as quadratic in terms of

approximately bosonic operators. This is resolved by decomposing a shell around the Fermi surface (the surface of the Fermi ball in momentum space) into patches  $B_\alpha$ ,  $\alpha = 1, \dots, M$ , as sketched in Fig. 1. If we then decompose the pair creation operators (up to a negligible contribution from the corridors between patches) as

$$b^*(k) = \sum_{\alpha=1}^M n_\alpha(k) b_\alpha^*(k),$$

$$b_\alpha^*(k) := \frac{1}{n_\alpha(k)} \sum_{\substack{p: p \in B_\alpha^c \cap B_x \\ p-k \in B_F \cap B_x}} a_p^* a_{p-k}^*$$

which allows for the linearization of the dispersion relation  $e(k)$ , we find



**Correlation Corrections as a Perturbation to the Quasi-free Approximation in Many-Body Quantum Systems, Fig. 1** Decomposition of the Fermi surface in momentum space  $2\pi\mathbb{Z}^3$  into  $M$  patches (Benedikter et al. 2020), where  $M = N^\alpha$  for some adjustable  $\alpha > 0$ . Patches are separated by thin corridors; patches extend radially inward and outward from the Fermi surface up to a distance corresponding to the largest momenta  $k$  for which  $\hat{V}(k)$  does not vanish (if necessary after imposing a cutoff). Patches are reflected by the origin to cover also the southern half of the sphere. (Sketch from Benedikter et al. (2020) under CC BY 4.0 license, <http://creativecommons.org/licenses/by/4.0/>, added  $\omega_\alpha$ )

$$[\mathbb{H}_0, b_\alpha^*(k)] = n_\alpha(k)^{-1} \sum_{\substack{p: p \in B_\alpha^c \cap B_x \\ p-k \in B_F \cap B_x}} (e(p) + e(p-k)) a_p^* a_{p-k}^* \simeq 2\hbar\kappa |k \cdot \hat{\omega}_\alpha| b_\alpha^*(k),$$

where  $\omega_\alpha$  is the vector pointing to the center of patch  $B_\alpha$  on the Fermi surface. This suggests to replace  $\mathbb{H}_0$  by a Hamiltonian of the form

$$\mathbb{D}_B := 2_\kappa \hbar \sum_{k \in \Gamma^{\text{nor}}} \sum_{\alpha=1}^M |k \cdot \hat{\omega}_\alpha| b_\alpha^*(k) b_\alpha(k),$$

on the grounds of it producing approximately the same commutator (using approximate CCR for the  $b_\alpha^*(k)$  and  $b_\alpha(k)$ ). By this heuristic argument, we have thus obtained an approximately quadratic bosonic Hamiltonian for the dominant term  $\mathbb{H}_0$  of Hartree–Fock theory and its dominant perturbation  $Q_B$ . With  $\mathbb{D}_B + Q_B$  quadratic in terms of approximately bosonic Hamiltonians, we may expect it to be approximately diagonalizable by a unitary operator mimicking a bosonic Bogoliubov transformation, i.e., taking the form of a quadratic exponential  $T$  in terms of the  $b_\alpha^*(k)$  and  $b_\beta(k)$ . This suggests to think of the ground state as being close to the form  $RT\Omega$ , where the presence of  $T$  (a quadratic exponential in the  $b$ - and  $b^*$ -operators, thus quartic in the fermionic operators) can be understood as the addition of a Bijl–Dingle–Jastrow–Mott factor to the Slater determinant  $R\Omega$ . In Benedikter et al. (2020), this trial state is used to prove the upper bound for Theorem 3.

Note that here two different (approximately) quasi-free theories appeared: the original fermionic quasi-free theory (the Hartree–Fock approximation) and the effective theory describing nontrivial fermionic correlations, which is approximately quadratic in terms of operators that satisfy approximate bosonic CCR.

Bosonization can also be used to study the dynamics of certain correlated initial data (Benedikter et al. 2022), leading to a Fock space norm approximation of the many-body Schrödinger equation. This can be seen as an improvement in precision over the approximation of the reduced

density matrices obtainable with the time-dependent Hartree–Fock equation (Benedikter et al. 2014, 2016).

## Future Directions

### Bosons

The natural direction following the study of scaling limits is the analysis of the thermodynamic limit, where many questions are still open; the most important and challenging being a proof of Bose–Einstein condensation, for now only available for scaling limits (Adhikari et al. 2021; Boccato et al. 2018, 2020; Brennecke et al. 2022; Lieb and Seiringer, 2002; Nam et al. 2016b, 2020; Seiringer 2011) and very specific models (Dyson et al. 1978; Kennedy et al. 1988). In the thermodynamic limit, one considers an unscaled Hamiltonian  $H_N$  of form (3) acting on  $L_s^2(\Lambda^N)$  and examines the limit  $N \rightarrow \infty$  and  $|\Lambda| \rightarrow \infty$  with fixed  $\rho = N/|\Lambda|$ .

Progress has recently been made on the ground state energy of the Bose gas in the thermodynamic limit. This is known as the *Lee–Huang–Yang formula*, first shown in Lee et al. (1957) by formal arguments. The Lee–Huang–Yang formula

$$\begin{aligned} e(\rho) &= \lim_{\substack{N, L \rightarrow \infty \\ \rho = N/|\Lambda|}} \frac{E(N, L)}{N} \\ &= 4\pi\rho a \left[ 1 + \frac{128}{15\sqrt{\pi}} (\rho a^3)^{1/2} + o\left((\rho a^3)^{1/2}\right) \right] \end{aligned} \quad (29)$$

is an expansion in small  $\rho a^3$  (diluteness condition) and is the analogue of (21). The factor  $\frac{128}{15\sqrt{\pi}}$  would be in fact the value of the second line in (21), if we replaced the sum with an integral. Upper bounds (Basti et al. 2021; Yau and Yin 2009) confirming (29) have been shown for interaction potentials  $V \in L^3(\mathbb{R}^3)$ , the lower bound confirming (29) for  $V \in L^3(\mathbb{R}^3)$  in Fournais and Solovej (2020). However, the formula is expected to be true also for hard spheres potentials, and the derivation in this case is still an open problem.

Only results for the leading order are available (Brietzke et al. 2020; Dyson 1957; Lieb and Yngvason 1998).

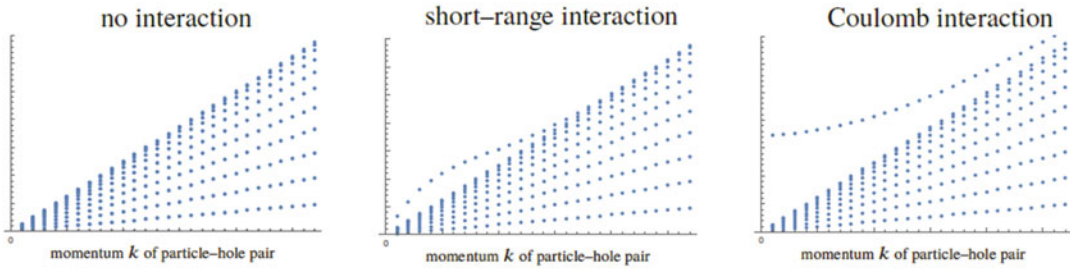
Generalizations of (21) and (29) to positive temperature have not yet been achieved, also in the Gross–Pitaevskii regime, and only the leading order has been studied (see (Deuchert and Seiringer 2020; Deuchert et al. 2019; Seiringer 2008; Yin 2010)). The works (Napiórkowski et al. 2018a, b), which we mentioned before, consider the Bogoliubov free energy functional, obtained by restricting the analysis to quasi-free states. The study shows that quasi-free states capture the free energy beyond leading order only in the case of weak interactions (where the scattering length can be approximated with its first Born approximation). For general interaction potentials, new methods are needed.

Only a few results are available in two space dimensions (Caraci et al. 2021; Lieb and Yngvason 2001); while an analogue of (17) still holds, a derivation of the second order is still an open problem. In one dimension, the Lieb–Liniger model for the delta-function Bose gas is exactly solvable and has been studied in Lieb (1963); Lieb and Liniger (1963).

More generally, the interest in Bose–Einstein condensates stems from several directions: condensates may find applications in quantum information processing (Byrnes et al. 2012, 2015), to simulate condensed matter systems (Bloch et al. 2008), and for precision measurements (Wang et al. 2005). The mathematically efficient description of entanglement effects, such as through the explicit trial states that we discussed here, may also support development in such more applied fields.

### Fermions

As in the bosonic case, a highly ambitious yet very important step is the analysis of the thermodynamic limit. In that case, one would like to the limit of large particle number  $N \rightarrow \infty$  and large volume  $\Lambda \rightarrow \infty$  while keeping the density  $\rho = N/|\Lambda|$  fixed. Only afterward one would consider the asymptotics of large density  $\rho \rightarrow \infty$ , in which the random phase approximation is expected to be valid



**Correlation Corrections as a Perturbation to the Quasi-free Approximation in Many-Body Quantum Systems, Fig. 2** Left: spectrum of the one-body problem (i.e., of the matrix  $\xi$  as defined by (14)) of the effective bosonic theory  $\mathbb{D}_B + Q_B$ . As the number of patches  $M \rightarrow \infty$ , a continuum is approximated. **Center:** If a short range interaction (no singularity of  $\hat{V}(k)$  for momenta

$k \rightarrow 0$ ) is present, the spectrum is only slightly deformed. **Right:** Even with the presence of a long-ranged interaction (behaving as  $\hat{V}(k) \sim |k|^{-2}$  for  $k \rightarrow 0$ ), almost the entire spectrum remains rigid, except for the plasmon mode at the top, responsible for screening of the interaction through collective oscillations

(in contrast, at low density one expects Wigner crystal behavior (Lewin et al. 2019) as the ground state). Of particular importance would be the treatment of the thermodynamic limit together with the Coulomb interaction  $\hat{V}(k) = |k|^{-2}$ , returning to the case which started the investigation in the early days of theoretical condensed matter physics. In this case, one expects to obtain at least the  $\rho \log \rho$ -term of the formula proposed by Gell-Mann and Brueckner (1957), where  $e(\rho)$  is the ground state energy in the thermodynamic limit, and  $e^{\text{HF}}(\rho)$  the Hartree–Fock minimum in the thermodynamic limit:

$$e(\rho) - e^{\text{HF}}(\rho) = C_{\text{MBP}} \rho \log \rho + C_{\text{GB}} \rho + o(\rho) \quad \text{for } \rho \rightarrow \infty.$$

A further important open question is the proof that the spectrum can be computed through bosonization, even for the finite volume case we discussed so far. The analysis of the effective correlation theory, as discussed in Benedikter (2021), gives rise to the spectrum of the effective theory as sketched in Fig. 2. In particular, with Coulomb interaction, the plasmon mode emerges, while the rest of the spectrum is only weakly deformed. The latter point would verify the mechanism proposed by Bohm and Pines (1953), according to which the plasmon mode causes the screening of the interaction, leaving only a weak remnant of the interaction potential. The idea that

electrons are thus converted into very weakly interacting quasiparticles is at the heart of Landau’s Fermi liquid theory, forming the basis for much of condensed matter theory. It is expected that bosonization may provide a microscopic explanation for Fermi liquid theory, as discussed by, e.g., (Castro Neto and Fradkin 1994a, b; Houghton et al. 2000; Houghton and Marston 1993).

More generally, quantum chemists recently use the random phase approximation to compute electron correlation effects and thus improve over density functional theory and Hartree–Fock computations, e.g., (Del Ben et al. 2015; Wilhelm et al. 2016). For example, it has been used to predict the structure of ice VI (Del Ben et al. 2014), one of the surprisingly many crystalline phases of water. Mathematically, it remains extremely hard to understand the validity of such approaches from the Schrödinger equation as a first principle. Nevertheless, we expect that the bosonization approach to the random phase approximation and the use of scaling limits to estimate the size of errors may still help to clarify the validity and the restrictions of such numerical computations.

## References

- Adhikari A, Brennecke C, Schlein B (2021) Bose–Einstein condensation beyond the Gross–Pitaevskii Regime. *Annales Henri Poincaré*. <https://doi.org/10/gmhss5>



- Bach V, Breteaux S, Chen T, Fröhlich J, Sigal IM (2016) The time-dependent Hartree-Fock-Bogoliubov equations for Bosons. arXiv:1602.05171 [math-ph]
- Basti G, Cenatiempo S, Schlein B (2021) A new second-order upper bound for the ground state energy of dilute Bose gases. *Forum of Mathematics, Sigma*, 9:E74. <https://doi.org/10.1017/fms.2021.66>
- Benedikter N (2021) Bosonic collective excitations in Fermi gases. *Rev Math Phys* 33(1):2060009. <https://doi.org/10/gjr35b>
- Benedikter N, Porta M, Schlein B (2014) Mean-field dynamics of fermions with relativistic dispersion. *J Math Phys* 55(2):021901. <https://doi.org/10/gdh7x8>
- Benedikter N, Porta M, Schlein B (2016) *Effective evolution equations from quantum dynamics* (SpringerBriefs in mathematical physics). Springer International Publishing
- Benedikter N, Sok J, Solovej JP (2018) The Dirac–Frenkel principle for reduced density matrices, and the Bogoliubov–de Gennes equations. *Ann Henri Poincaré* 19(4):1167–1214. <https://doi.org/10/gc7p9m>
- Benedikter N, Nam PT, Porta M, Schlein B, Seiringer R (2020) Optimal upper bound for the correlation energy of a fermi gas in the Mean-Field Regime. *Commun Math Phys* 374(3):2097–2150. <https://doi.org/10/gf7fkt>
- Benedikter N, Nam PT, Porta M, Schlein B, Seiringer R (2021a) Correlation energy of a weakly interacting Fermi gas. *Invent Math* 225(3):885–979. <https://doi.org/10/gnkrwz>
- Benedikter N, Porta M, Schlein B, Seiringer R (2021b) Correlation energy of a weakly interacting fermi gas with large interaction potential. arxiv:2106.13185 [cond-mat, physics:math-ph] (2021)
- Benedikter N, Nam PT, Porta M, Schlein B, Seiringer R (2022) Bosonization of Fermionic Many-Body Dynamics. *Annales Henri Poincaré* 23:1725–1764. <https://doi.org/10.1007/s00023-021-01136-y>
- Bloch I, Dalibard J, Zwerger W (2008) Many-body physics with ultracold gases. *Rev Mod Phys* 80(3):885–964. <https://doi.org/10.1103/RevModPhys.80.885>
- Boccato C (2021) The excitation spectrum of the Bose gas in the Gross–Pitaevskii regime. *Rev Math Phys* 33(01):2060006. <https://doi.org/10.1142/S0129055X20600065>
- Boccato C, Brennecke C, Cenatiempo S, Schlein B (2018) Complete Bose–Einstein condensation in the Gross–Pitaevskii regime. *Commun Math Phys* 359(3):975–1026. <https://doi.org/10.1007/s00220-017-3016-5>
- Boccato C, Brennecke C, Cenatiempo S, Schlein B (2019) Bogoliubov theory in the Gross–Pitaevskii limit. *Acta Math* 222(2):219–335. <https://doi.org/10.4310/ACTA.2019.v222.n2.a1>
- Boccato C, Brennecke C, Cenatiempo S, Schlein B (2020) Optimal rate for Bose–Einstein condensation in the Gross–Pitaevskii Regime. *Commun Math Phys* 376(2):1311–1395. <https://doi.org/10/ggk8hq>
- Bogolubov NN (1947) On the theory of Superfluidity. *J Phys (USSR)* 11(23–32). <https://doi.org/10.1016/B978-0-08-015816-7.50020-1>
- Bohm D, Pines D (1953) A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas. *Phys Rev* 92(3):609–625. <https://doi.org/10/fkm4kf>
- Boßmann L, Petrat S, Seiringer R (2021) Asymptotic expansion of low-energy excitations for weakly interacting bosons. *Forum Math Sigma* 9:e28. <https://doi.org/10.1017/fms.2021.22>
- Brennecke C, Schlein B, Schraven S (2022) Bose–Einstein condensation with optimal rate for trapped Bosons in the Gross–Pitaevskii Regime. *Math Phys Anal Geom* 25(12). <https://doi.org/10.1007/s11040-022-09424-7>
- Brietzke B, Fournais S, Solovej JP (2020) A simple 2nd order lower bound to the energy of dilute Bose gases. *Communications in Mathematical Physics* 376(1):323–351. arXiv:1901.00539 [math-ph]
- Byrnes T, Wen K, Yamamoto Y (2012) Macroscopic quantum computation using Bose-Einstein condensates. *Phys Rev A* 85(4):040306(R). <https://doi.org/10/ghtzvm>
- Byrnes, T., Rosseau, D., Khosla, M., Pyrkov, A., Thomasen, A., Mukai, T., Koyama, S., Abdelrahman, A., Ilo-Okeke, E.: Macroscopic quantum information processing using spin coherent states. *Optics Communications* 337, 102–109 (2015). <https://doi.org/10/gnnbbf>
- Caraci C, Cenatiempo S, Schlein B (2021) Bose-Einstein condensation for two dimensional bosons in the Gross-Pitaevskii regime. *Journal of Statistical Physics* 183(3):1–72. arXiv:2011.05962 [cond-mat, physics:math-ph]
- Castro Neto AH, Fradkin E (1994a) Bosonization of Fermi liquids. *Phys Rev B* 49(16):10877–10892. <https://doi.org/10/cxjb5q>
- Castro Neto AH, Fradkin E (1994b) Bosonization of the low energy excitations of Fermi liquids. *Phys Rev Lett* 72(10):1393–1397. <https://doi.org/10.1103/PhysRevLett.72.1393>
- Cenatiempo S (2019) Bogoliubov theory for dilute Bose gases: The Gross-Pitaevskii regime. *J Math Phys* 60(8):081901. <https://doi.org/10/gmh3bp>
- Christiansen MR, Hainzl C, Nam PT (2021) The random phase approximation for interacting fermi gases in the Mean-Field Regime. arXiv:2106.11161 [cond-mat, physics:math-ph]
- Del Ben M, VandeVondele J, Slater B (2014) Periodic MP2, RPA, and boundary condition assessment of hydrogen ordering in ice XV. *J Phys Chem Lett* 5(23):4122–4128. <https://doi.org/10/f6r6t3>
- Del Ben M, Schütt O, Wentz T, Messmer P, Hutter J, VandeVondele J (2015) Enabling simulation at the fifth rung of DFT: large scale RPA calculations with excellent time to solution. *Comput Phys Commun* 187:120–129. <https://doi.org/10/f6t3nf>
- Deuchert A, Seiringer R (2020) Gross–Pitaevskii limit of a homogeneous Bose gas at positive temperature. *Arch Ration Mech Anal* 236(3):1217–1271. <https://doi.org/10/gmhrpz>
- Deuchert A, Seiringer R, Yngvason J (2019) Bose–Einstein condensation in a dilute, trapped gas at positive

- temperature. *Commun Math Phys* 368(2):723–776. <https://doi.org/10/gfsz3t>
- Dyson FJ (1957) Ground-state energy of a hard-sphere gas. *Phys Rev* 106(1):20–26. <https://doi.org/10/bnctfg>
- Dyson FJ, Lieb EH, Simon B (1978) Phase transitions in quantum spin systems with isotropic and nonisotropic interactions. *J Stat Phys* 18(4):335–383. <https://doi.org/10/bn3d6w>
- Erdos L, Schlein B, Yau HT (2008) Ground-state energy of a low-density Bose gas: A second order upper bound. *Phys Rev A* 78(5) <https://doi.org/10/cz48kn>
- Falconi M, Giacomelli EL, Hainzl C, Porta M (2021) The dilute Fermi gas via Bogoliubov theory. *Ann Henri Poincaré* 22(7):2283–2353. <https://doi.org/10/gmhrtg>
- Fournais S, Solovej JP (2020) The energy of dilute Bose gases. *Ann Math* 192(3):893–976. <https://doi.org/10/gh6ht7>
- Gell-Mann M, Brueckner KA (1957) Correlation energy of an electron gas at high density. *Phys Rev* 106(2):364–368. <https://doi.org/10.1103/PhysRev.106.364>
- Gontier D, Lewin M (2019) Spin symmetry breaking in the translation-invariant Hartree-Fock Uniform Electron Gas. *SIAM Journal on Mathematical Analysis* 51(4):3388–3423. arXiv:1812.07679 [cond-mat, physics:math-ph]
- Gontier D, Hainzl C, Lewin M (2019) Lower bound on the Hartree-Fock energy of the electron gas. *Phys Rev A* 99(5):052501. <https://doi.org/10/gt2fp8>
- Graf GM, Solovej JP (1994) A correlation estimate with applications to quantum systems with coulomb interactions. *Rev Math Phys* 6(05a):977–997. <https://doi.org/10/bw8d5g>
- Hainzl C, Seiringer R (2016) The Bardeen–Cooper–Schrieffer functional of superconductivity and its mathematical properties. *J Math Phys* 57(2):021101. <https://doi.org/10/f8c9gt>
- Hainzl C, Porta M, Rexze F (2020) On the correlation energy of interacting fermionic systems in the Mean-Field Regime. *Commun Math Phys* 374(2):485–524. <https://doi.org/10/ggkwjnm>
- Heisenberg W (1947) Zur Theorie der Supraleitung. *Zeitschrift für Naturforschung A* 2(4):185–201. <https://doi.org/10.1515/zna-1947-0401>
- Houghton A, Marston JB (1993) Bosonization and fermion liquids in dimensions greater than one. *Phys Rev B* 48(11):7790–7808. <https://doi.org/10/dnt25d>
- Houghton A, Kwon HJ, Marston JB (2000) Multi-dimensional bosonization. *Adv Phys* 49(2):141–228. <https://doi.org/10/csxtkj>
- Kennedy T, Lieb EH, Shastry BS (1988) The XY model has long-range order for all spins and all dimensions greater than one. *Phys Rev Lett* 61(22):2582–2584. <https://doi.org/10/cc46zj>
- Landau L (1941) Theory of the superfluidity of Helium II. *Phys Rev* 60(4):356–358. <https://doi.org/10/b84r26>
- Lee TD, Huang K, Yang CN (1957) Eigenvalues and Eigenfunctions of a Bose system of hard spheres and its low-temperature properties. *Phys Rev* 106(6):1135–1145. <https://doi.org/10/bkrhf8>
- Lewin M, Lieb EH, Seiringer R (2019) Floating Wigner crystal with no boundary charge fluctuations. *Phys Rev B* 100(3):035127. <https://doi.org/10/gmhrrs9>
- Lieb EH (1963) Exact analysis of an interacting Bose Gas. II. The excitation spectrum. *Phys Rev* 130(4):1616–1624. <https://doi.org/10/b2n7sm>
- Lieb EH, Liniger W (1963) Exact analysis of an interacting Bose Gas. I. The general solution and the ground state. *Phys Rev* 130(4):1605–1616. <https://doi.org/10/dhgb9j>
- Lieb EH, Seiringer R (2002) Proof of Bose-Einstein condensation for dilute trapped gases. *Phys Rev Lett* 88(17) <https://doi.org/10/fbqzsm>
- Lieb EH, Yngvason J (1998) Ground state energy of the low density bose gas. *Phys Rev Lett* 80(12):2504–2507. <https://doi.org/10/cx23qh>
- Lieb EH, Yngvason J (2001) The ground state energy of a dilute two-dimensional bose gas. *J Stat Phys* 103(3):509–526. <https://doi.org/10/cmsbxm>
- Lieb EH, Seiringer R, Yngvason J (2000) Bosons in a trap: a rigorous derivation of the Gross-Pitaevskii energy functional. *Phys Rev A* 61(4):043602. <https://doi.org/10/cdwvnb>
- Lubich C (2008) From quantum to classical molecular dynamics: reduced models and numerical analysis. Zurich lectures in advanced mathematics. European Mathematical Society, Zürich. <https://doi.org/10.4171/067>
- Macke W (1950) über die Wechselwirkungen im Fermi-Gas. Polarisationerscheinungen, Correlationsenergie, Elektronenkondensation. *Zeitschrift für Naturforschung A* 5(4):192–208
- Nam PT, Napiórkowski M, Solovej JP (2016a) Diagonalization of bosonic quadratic Hamiltonians by Bogoliubov transformations. *J Funct Anal* 270(11):4340–4368. <https://doi.org/10/f8kchm>
- Nam PT, Rougerie N, Seiringer R (2016b) Ground states of large bosonic systems: The Gross–Pitaevskii limit revisited. *Anal PDE* 9(2):459–485. <https://doi.org/10/f8sfpq>
- Nam PT, Napiórkowski M, Ricaud J, Triay A (2020) Optimal rate of condensation for trapped bosons in the Gross–Pitaevskii regime. *Analysis & PDE, Mathematical Sciences Publishers*, In press. arXiv:2001.04364 [math-ph]
- Napiórkowski M (2021) Dynamics of interacting bosons: A compact review. *Proceedings of the Workshop on Density Functionals for Many-Particle Systems*, September 2019, Singapore. arXiv:2101.04594 [cond-mat, physics:math-ph]
- Napiórkowski M, Reuvers R, Solovej JP (2018a) The Bogoliubov free energy functional I: existence of minimizers and phase diagram. *Arch Ration Mech Anal* 229(3):1037–1090. <https://doi.org/10/czq9>
- Napiórkowski M, Reuvers R, Solovej JP (2018b) The Bogoliubov free energy functional II: the dilute limit. *Commun Math Phys* 360(1):347–403. <https://doi.org/10/gdhz65>
- Reed M, Simon B (2012) *Methods of Modern mathematical physics, volume 4: analysis of operators*. Academic Press, San Diego, California

- Rougerie N (2021) Scaling limits of bosonic ground states, from many-body to nonlinear Schrödinger. EMS Surveys in Mathematical Sciences, EMS, In press. arXiv:2002.02678 [cond-mat, physics:math-ph]
- Sawada K (1957) Correlation energy of an electron gas at high density. *Phys Rev* 106(2):372–383. <https://doi.org/10.1103/PhysRev.106.372>
- Sawada K, Brueckner KA, Fukuda N, Brout R (1957) Correlation energy of an electron gas at high density: plasma oscillations. *Phys Rev* 108(3):507–514. <https://doi.org/10.1103/PhysRev.108.507>
- Seiringer R (2008) Free energy of a dilute bose gas: lower bound. *Commun Math Phys* 279(3):595–636. <https://doi.org/10/dsffnh>
- Seiringer R (2011) The excitation spectrum for weakly interacting Bosons. *Commun Math Phys* 306(2): 565–578. <https://doi.org/10/d3f5zp>
- Solovej JP (2014) Many body quantum mechanics. Lecture Notes Erwin Schrödinger Institute Vienna. <http://web.math.ku.dk/~solovej/MANYBODY/>
- Spohn H (1980) Kinetic equations from Hamiltonian dynamics: Markovian limits. *Rev Mod Phys* 52(3): 569–615. <https://doi.org/10/c2z38s>
- Wang YJ, Anderson DZ, Bright VM, Cornell EA, Diot Q, Kishimoto T, Prentiss M, Saravanan RA, Segal SR, Wu S (2005) Atom Michelson interferometer on a chip using a Bose-Einstein condensate. *Phys Rev Lett* 94(9):090405. <https://doi.org/10/dxxpkm>
- Wigner E (1934) On the interaction of electrons in metals. *Phys Rev* 46(11):1002–1011. <https://doi.org/10/d36jvg>
- Wilhelm J, Seewald P, Del Ben M, Hutter J (2016) Large-scale cubic-scaling random phase approximation correlation energy calculations using a Gaussian basis. *J Chem Theory Comput* 12(12):5851–5859. <https://doi.org/10/f9ghm2>
- Williamson J (1936) On the algebraic problem concerning the normal forms of linear dynamical systems. *Am J Math* 58(1):141–163. <https://doi.org/10/fjn72s>
- Yau HT, Yin J (2009) The second order upper bound for the ground energy of a bose gas. *J Stat Phys* 136(3): 453–503. <https://doi.org/10/c8wns4>
- Yin J (2010) Free energies of dilute bose gases: upper bound. *J Stat Phys* 141(4):683–726. <https://doi.org/10/cppnzh>

# Perturbation of Equilibria in the Mathematical Theory of Evolution

Angel Sánchez

Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Madrid, Spain  
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Zaragoza, Spain

## Article Outline

Glossary

Definition of the Subject

Introduction

Evolution on a Fitness Landscape

Stability of Equilibria on a Fitness Landscape

Perturbation of Equilibria on a Fitness Landscape

Frequency Dependent Fitness: Game Theory

Equilibria in Evolutionary Game Theory

Perturbations of Equilibria in Evolutionary Game Theory

Spatial Perturbations

Time Scales

Future Directions

Bibliography

## Glossary

**Evolutionarily stable equilibria (ESS)** An ESS is a set of frequencies of different types of individuals in a population that can not be invaded by the evolution of a single mutant. It is the evolutionary counterpart of a Nash equilibrium.

**Fitness landscape** A metaphorical description of fitness as a function of individual's genotypes or phenotypes in terms of a multivariable function that does not depend on any external influence.

**Genetic locus** The position of a gene on a chromosome. The different variants of the gene that can be found at the same locus are called alleles.

**Nash equilibrium** In classical game theory, a Nash equilibrium is a set of strategies, one for each player of the game, such that none of them can improve her benefits by unilateral changes of strategy.

**Scale free network** A graph or network such that the degrees of the nodes are taken from a power-law distribution. As a consequence, there is not a typical degree in the graph, i.e., there are no typical scales.

**Small-world network** A graph or network of  $N$  nodes such that the mean distance between nodes scales as  $\log N$ . It corresponds to the well-known "six degrees of separation" phenomenon.

## Definition of the Subject

The importance of evolution can hardly be overstated. As the Jesuit priest Pierre Teilhard de Chardin put it,

Evolution is a general postulate to which all theories, all hypotheses, all systems must hence forward bow and which they must satisfy in order to be thinkable and true. Evolution is a light which illuminates all facts, a trajectory which all lines of thought must follow – this is what evolution is.

Darwin's evolution theory is based on three fundamental principles: reproduction, mutation and selection, which describe how populations change overtime and how new forms evolve out of old ones. Starting with W. F. R. Weldon, whom at the beginning of the twentieth century realized that "the problem of animal evolution is essentially a statistical problem", and blooming in the 30's with Fisher, Haldane and Wright, numerous mathematical descriptions of the resulting evolutionary dynamics have been proposed, developed and studied. Deeply engraved in these frameworks are the mathematical concepts of equilibrium and

stability, as descriptions of the observed population compositions and their lifetimes. Many results have been obtained regarding the stability of equilibria of evolutionary dynamics in idealized circumstances, such as infinite populations or global interactions. In the evolutionary context, stability is peculiar, in the sense that it is entangled with collective effects arising from the interaction of individuals. Therefore, perturbations of the idealized mathematical framework representing more realistic situations are of crucial importance to understand stability of equilibria.

## Introduction

The idea of evolution is a simple one: Descent with modification acted upon by natural selection. Descent with modification means that we consider a population of replicators, entities capable of reproducing themselves, in which reproduction is not exact and allows for small differences between parents and offspring. Natural selection means that different entities reproduce in different quantities because their abilities are also different: some are more resistant to external factors, some need less resources, some simply reproduce more... While in biology these replicators are, of course, living beings, the basic ingredients of evolution have by now transcended the realm of biology into the kingdom of objects such as computer codes (thus giving rise, e.g., to genetic algorithms).

The process of evolution is often described in short as “the survival of the fittest”, meaning that those replicators that succeed and thrive are more fit than those which progressively disappear. This statement is not very appropriate, because evolution does not imply making organisms or entities more fit; it is simply a consequence of differential reproduction in the face of selection pressure. On the other hand, it leads to a tautology: The question of which are the fittest organisms is answered by saying that they are those that survive. It is then clear that a correct use of the concept of fitness is at the crux of any attempt to formalize mathematical evolution theory.

In this article we are going to discuss two manners to deal mathematically with the concept of fitness. The first and simplest one is to resort to a

fitness landscape, whose basic feature is that the fitness of a given individual depends only on the individual’s characteristics and not on external factors. We will present this approach in sections “[Evolution on a Fitness Landscape](#)” and “[Stability of Equilibria on a Fitness Landscape](#)” below, to subsequently discuss the effect of perturbations on the equilibria described by this picture in section “[Perturbation of Equilibria on a Fitness Landscape](#)”. To go beyond the fitness landscape picture one has to introduce frequency-dependent selection, i.e., to remove the independence of the fitness from external influences. In sections “[Frequency Dependent Fitness: Game Theory](#)” and “[Equilibria in Evolutionary Game Theory](#)” we consider the evolutionary game theory approach to this way to model evolution and, as before, analyze the perturbations of its equilibria in section “[Perturbations of Equilibria in Evolutionary Game Theory](#)”. Section “[Future Directions](#)” summarizes the questions that remain open in this field.

## Evolution on a Fitness Landscape

The metaphor of evolution on a “fitness landscape” reaches back at least to (Wright 1932): Drawing on the connection between fitness and adaptation, fitness is defined as the expected number of offspring of a given individual that reach adulthood, and thus represents a measure of its adaptation to the environment. In this context, fitness landscapes are used to visualize the relationship between genotypes (or phenotypes) and reproductive success. It is assumed that every genotype has a well defined fitness, in the sense above, and that this fitness is the “height” of the landscape. Genotypes which are very similar are said to be “close” to each other, while those that are very different are “far” from each other. The two concepts of height and distance are sufficient to form the concept of a “landscape”. The set of all possible genotypes, their degree of similarity, and their related fitness values is then called a fitness landscape.

Fitness landscapes are often conceived of as ranges of mountains. There exist local peaks (points from which all paths are downhill, i.e. to lower fitness) and valleys (regions from which most paths lead uphill). A fitness landscape with



many local peaks surrounded by deep valleys is called rugged. If all genotypes have the same replication rate, on the other hand, a fitness landscape is said to be flat. A sketch of such a fitness landscape, showing the dependence of the fitness on two different “characteristics” or “genes”, is shown in Fig. 1. Of course, the true fitness landscape would need a highly multidimensional space for its representation, as it would depend on all the characteristics of the organism, even those it still does not show. Therefore, the sketch is an extreme oversimplification, only to suggest the structure of a rugged fitness landscape.

Given the immense complexity of the genotype-fitness mapping, theoretical models have to make a variety of simplifying assumptions. Most models in biological literature focus on the effect of one or a few genetic loci on the fitness of individuals in a population, assuming that each of the considered loci can be occupied by a limited number of different alleles that have different effects on the fitness, and that the rest of the genome is part of the invariant environment. This approximation, the first attempt to obtain analytical results for changes in the gene pool of a population under the influence of inheritance, selection and mutation is the pioneering work of Fisher, Haldane and Wright, who founded the field of population genetics. Their method of randomly drawing the genes of the daughter population from the pool of parent genes, with weights proportional to the fitness, proved to be very successful at calculating the evolution of allele frequencies from one generation to the next, or the chances of a new mutation to spread through a population,

even taking into account various patterns of mating, dominance effects, nonlinear effects between different genes, etc. Population genetics has since then developed into a mature field with a sophisticated mathematical apparatus, and with wide-ranging applications.

## Stability of Equilibria on a Fitness Landscape

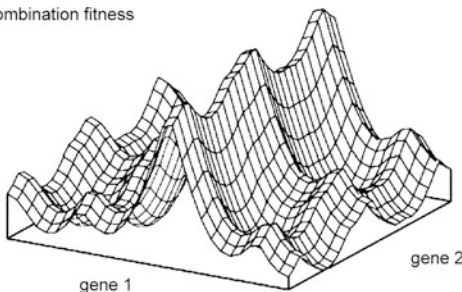
The simplified pictures we have just described lead to a description in terms of dynamical systems, and therefore the stability of its equilibria can be studied by means of standard techniques. In principle, one can envisage the evolution of a population on a fitness landscape in the usual frame of particle dynamics on a potential. Every individual is a point in the space of phenotypes or genotypes, and evolves towards the maxima of the fitness; in the potential picture, the potential is given by minus the fitness. Furthermore, the dynamics is overdamped, i.e., there are no oscillations around the equilibria. Maxima of the fitness are therefore the equilibria of the evolutionary process. That this is so is a consequence of Fisher’s theorem (Fisher 1958), whose original derivation is very general but quite complicated. Following (Drossel 2001; Gintis 2000), we prefer here to present two simpler situations: an asexual population, and a sexually reproducing population where the fitness is determined by a single gene with two alleles.

For an asexually reproducing population, the derivation of Fisher’s theorem is straightforward: Let  $y_i$  be the number of genes  $i$  in the population, and  $y$  the total number of genes; then  $p_i \equiv y_i/y$  is the frequency of genotype  $i$  in the population. If  $W_i$  is gene  $i$ ’s fitness, sticking to the interpretation of fitness in terms of offspring, the number of individuals carrying gene  $i$  in the next generation is  $W_i y_i$  and, subsequently, the change in the frequency  $p_i$  from one generation to the next is

$$\Delta p_i = p_i (W_i - \bar{W}) / \bar{W},$$

leading to a change in mean fitness

gene combination fitness



**Perturbation of Equilibria in the Mathematical Theory of Evolution, Fig. 1** Sketch of a fitness landscape



$$\frac{\Delta \bar{W}}{\bar{W}} = \frac{\sum_i w_i \Delta p_i}{\bar{W}} = \frac{\sum_i p_i (w_i^2 - \bar{W}^2)}{\bar{W}^2}$$

which is proportional to the genetic variance in fitness. If the fitness changes from one generation to the next are small, this becomes an equation which states that the rate of change in fitness is identical to the genetic variance in fitness.

When reproduction is sexual, we note that  $p_i$  is the frequency of gene  $i$ . Assuming for simplicity that only two alleles are possible at the genetic locus of interest, the fitness of type 1 is  $w_{11}p_1 + w_{12}p_2$ , where  $w_{ij}$  is the fitness of an individual carrying alleles  $i$  and  $j$ , and hence the number of 1 genes will be  $y_1(w_{11}p_1 + w_{12}p_2)$ . We then have the differential equation

$$\dot{y}_1 = y_1(w_{11}p_1 + w_{12}p_2). \quad (1)$$

It is enough then to differentiate the identity  $\ln p_1 = \ln y_1 - \ln y$  and use a little algebra to show that  $y_1$  obeys the replicator equation:

$$\dot{p}_1 = p_1(w_1 - \bar{w}) \quad (2)$$

Fisher's theorem then states that fitness increases along trajectories of this equation: Indeed, by noting that the average fitness is

$$\bar{w} = \sum_{i,j=1,2} w_{ij}p_i p_j, \quad (3)$$

differentiating and using the replicator equation we arrive at the final result

$$\dot{\bar{w}} = 2 \sum_i p_i (w_i - \bar{w})^2. \quad (4)$$

Fisher's theorem thus means that an evolving population will typically climb uphill in the fitness landscape, by a series of small genetic changes, until a local optimum is reached. This is due to the fact that the average fitness of the population always increases, as we have just shown; hence the analogy with overdamped dynamics on a (inverted) potential function. Furthermore, because of this result, the population remains there, at the

equilibrium point, because it cannot reduce its fitness. We then realize that all equilibria in a fitness landscape within the interpretation of fitness as the reproductive success are stable.

## Perturbation of Equilibria on a Fitness Landscape

In order to understand the possible breakdowns of stability in the fitness landscape picture, one has to look carefully at the hypothesis of Fisher's theorem. We have not stated it in a formal manner, hence it is important to summarize here the main ones:

- Population is infinite.
- There are no mutations (i.e., the only source of every gene or species is reproduction).
- There is only one population (i.e., there are no population fluxes or migrations between separate groups).
- Fitness depends only on the individual's genotype and not on the other individuals.

It is then clear that, even if it is mathematically true, the applicability of Fisher's theorem is a completely different story, and as a consequence, the conclusion that maxima of the fitness landscape are stable may be wrong when discussing real systems. A detailed discussion of all these issues can be found in (Drossel 2001), and we refer the reader to her paper for a thorough discussion of all these factors. For our present purposes, namely to show that these perturbations can change the stability of the equilibria, it will suffice to present a few ideas about the case of finite population size. Afterwards, the rest of the paper will proceed along the idea of fitness depending on other individuals, giving up the paradigm of a fixed fitness landscape.

The subject of finite size populations is the subject of fluctuations and its main consequences, genetic drift and stochastic escape. Regarding the first concept, as compared to natural selection, i.e., to the tendency of beneficial alleles to become more common over time (and detrimental ones less common), genetic drift is the

fundamental tendency of any allele to vary randomly in frequency over time due to statistical variation alone, so long as it does not comprise all or none of the distribution. In other words, even when individuals face the same odds, they will differ in their success. A rare succession of chance events can thus bring a trait to predominance, causing a population or species to evolve (in fact, this idea is at the core of the neutral theory of evolution, first proposed by (Kimura 1983)). On the other hand, stochastic escape refers to the situation in which a population of individuals placed at a maximum of the fitness landscape may leave this maximum due to fluctuations. Obviously, both genetic drift and stochastic escape affect the stability of the maxima as predicted by Fisher's theorem.

One consequence of finite population sizes and fluctuations in the composition of a population is that genes get lost from the gene pool. If there is no new genetic input through mutation or migration, the genetic variability within a population decreases with time. After sufficiently many generations, all individuals will carry the same allele of a given gene. This allele is said to have become fixed. In the absence of selection, the probability that a given allele will become fixed is proportional to the number of copies in the initial population. Thus, if a new mutant arises that has no selective advantage or disadvantage, this mutant will spread through the entire population with a probability  $1/M$ ,  $M$  being the population size. If the individuals of the population are diploid, each carries two sets of genes, and  $M$  must be taken as the number of sets of genes, i.e., as twice the population size. On the other hand, it can be shown (Drossel 2001) that the probability that a mutant that conveys a small fitness increase by a factor  $1 + s$  has a probability of the order  $s$  to spread through a population. In populations of sizes much smaller than  $1/s$ , this selective advantage is not felt, because mutations that carry no advantage become fixed at a similar rate. In the same manner, a mutation that decreases the fitness of its carrier by a factor  $1 - s$ , is not felt in a population much smaller than  $1/s$ . An interesting consequence of these results is that the rate of neutral (or effectively neutral) substitutions is

independent of the population size. The reason is that the probability that a new mutant is generated in the population is proportional to  $M$ , while its probability of becoming fixed is  $1/M$ .

## Frequency Dependent Fitness: Game Theory

In the preceding sections we have considered the case when the fitness depends on the genotype, but is independent of the composition of the population, i. e., the presence of individuals of the same genotype or of other genotypes does not change the fitness of the focal one. This assumption, that allows for an intuitive picture in terms of a fitness landscape, is clearly an over-simplification, as was already mentioned above. For instance, consider an homogeneous population in a closed environment. The population will grow at a pace given by the fitness of its individuals until it eventually exhausts the available resources or even physically fills the environment. Therefore, even if the individuals are all equal, their fitness will not be the same if there are only a few of them or if there are very many. Another trivial example is the effect on the fitness of the presence or absence of predators of the species of interest; clearly, predators will reduce the fitness (understood as above in a reproductive sense) of their prey.

Therefore, individuals will evolve subject not only to external influences but also to their mutual competition, both intra-specific and inter-specific. This leads us to consider frequency-dependent selection, which can be described by very many, different theoretical approaches. These include game theory as well as discrete and continuous genetic models, and the concepts of kin selection, group selection, and sexual selection. Among the possible dynamical patterns arising, there are single fixed points, lines of fixed points, runaway, limit cycles, and chaos. A review of all these descriptions, whose use to model evolution depends on the specific issues one is interested in, is clearly far beyond the scope of this article and, hence, we have focused on evolutionary game theory as a particularly suited case study to show the effects of perturbations on equilibria.

Before going into the study of evolutionary game theory, we need to summarize briefly a few key concepts about its originating theory, namely (classical) game theory. Pioneered in the early XIX century by the economist Cournot, game theory was introduced by the brilliant, multifaceted mathematician John von Neumann in 1928, and it was first presented as a specific subject in von Neumann's book (with Oskar Morgenstern) *Theory of Games and Economic Behavior* in 1944. Since then, game theory has been used to model strategic situations, i. e., situations in which actors or agents follow different strategies (meaning that they choose among different possible actions or behaviors) to maximize their benefit, usually referred to as *payoff*. These arise in very many different contexts, from biology and psychology to philosophy through politics, economics or computer science.

The central concept of game theory is the Nash equilibrium (Nash 1950), introduced by the mathematician John Nash in 1955, awarded with a Nobel Prize in Economics almost 40 years later for this work. A set of strategies, one for each participant in the game, is a Nash equilibrium if every strategy is the best response (in terms of maximizing the player's payoff) to the subset of the strategies of the rest of the players. In this case, if all players use strategies belonging to a Nash equilibrium, none of them will have any incentive to change her behavior. In this situation we indeed have the equivalent of the traditional concept of equilibrium in dynamical systems: players keep playing the same strategy as, given the behavior of the others, they follow the optimal strategy (note that this does not mean the strategy is optimal in absolute terms: it is only optimal in view of the actions of the rest).

## Equilibria in Evolutionary Game Theory

In the seventies, game theory, which as proposed by von Neumann and Nash was to be used to understand economic behavior, entered the realm of biology through the pioneering work of John Maynard-Smith and George Price (Maynard-Smith and Price 1973), who introduced the evolutionary

version of the theory. The key contribution of their work was a new interpretation of the general framework of game theory in terms of populations instead of individual players. While traditional game theoretical players behaved following some strategy and could change it to improve their performance, in the picture of Maynard-Smith and Price individuals had a fixed strategy, determined by their genotype, and different strategies were represented by sub-populations of individuals. In this representation, changes of strategies correspond to the replacement of the individuals by their offspring, possibly with mutations. Payoffs obtained by individuals in the game are accordingly understood as fitness, the reproductive rate that governs how the replacement occurs.

There is a large degree of arbitrariness as to the evolutionary dynamics of the populations. All we have said so far is that fitness, obtained through the game, determines the composition of the population at the next time step (or instant, if we think of continuous time). Probably the most popular choice (but by no means the only one, see (Page and Nowak 2002) for different evolutionary proposals and their relationships, see also (Hofbauer and Sigmund 2003) for other dynamics) is to use the replicator equation we have previously found to describe the evolution of the frequency  $y_i$  of strategists of type  $i$ :

$$\dot{y}_i = y_i(w_i - \bar{w}). \quad (5)$$

It is important to note that the steps leading to the derivation of this equation are the same as above, and therefore for it to be applicable in principle one must keep in mind the same hypotheses. The difference is that now the fitness is not a constant but rather it is determined by a game, which enters the equation in the following way.

Let us call  $\mathbf{A}$  the payoff matrix of the game (for simplicity, we will consider only symmetric games), whose entries  $a_{ij}$  are the payoffs to an individual using strategy  $i$  facing another using strategy  $j$ . Assuming the frequencies  $y_i(t)$  are differentiable functions, if individuals meet randomly and then engage in the game, and this takes place very many (infinite) times, then  $(\mathbf{A}\mathbf{y})_i$  is the expected payoff for type  $i$  individuals in a

population described by the vector  $\mathbf{y}$ , whose components are the frequencies of each type. By the same token, the average payoff in the population is  $\bar{w} = \mathbf{y}^T \mathbf{A} \mathbf{y}$ , so substituting in (5) we are left with

$$\dot{y}_i = y_i [(\mathbf{A} \mathbf{y})_i - \mathbf{y}^T \mathbf{A} \mathbf{y}], \quad (6)$$

where we now see explicitly how the game affects the evolution. Nevertheless, it is also clear that this rule is arbitrary, and there are many other options one can use to postulate how the population evolves. We will come back to this issue when considering perturbations of the equilibria.

If Nash equilibrium is the key concept in game theory, evolutionarily stable strategy is the relevant one in its evolutionary counterpart (Maynard-Smith 1982). defined evolutionarily stable strategy (ESS) as a strategy such that, if every individual in the population uses it, no other (mutant) strategy could invade the population by natural selection. It is trivial to show that, in terms of the payoffs of the game, for strategy  $i$  to be an ESS, one of the following two conditions must hold:

$$a_{ii} > a_{ij} \forall j \neq i, \text{ or} \quad (7)$$

$$a_{ii} = a_{ij} \text{ for some } j \text{ and } a_{ij} > a_{ji} \quad (8)$$

If the first condition is fulfilled, we speak of *strict* ESS. It is important to realize that this concept is absolutely general and, in particular, it does not depend on the evolutionary dynamics of choice (in so far as it favors the strategies that receive the best payoffs).

Of course, the two concepts, Nash equilibrium and ESS, are related. This is in fact one of the reasons why evolutionary game theory ended up appealing to the economists, who faced the question as to how individuals ever get to play the Nash equilibrium strategies: They now had a dynamical way that might precisely describe that process and, furthermore, to decide which Nash equilibrium was selected if there were more than one. To show the connection, one must decide on a dynamical rule, for which we will stay within the framework of the replicator dynamics. For this specific evolutionary dynamics, it can be rigorously

shown that (see, e. g., (Hendry and Kinnison 1999; Hofbauer and Sigmund 2003))

1. if  $y_0$  is a Nash equilibrium, it is a rest point (a zero of the rhs of (5);
2. if  $y_0$  is a strict Nash equilibrium, it is asymptotically stable;
3. if  $y_0$  is a rest point and is the limit of an interior orbit for  $t \rightarrow \infty$ , then it is a Nash equilibrium; and
4. if  $y_0$  is a stable rest point, it is a Nash equilibrium.

This means that there indeed is a relationship between Nash equilibria and ESS, but more subtle than could appear at first. Probably, the most important non trivial aspect of this result is that not all ESS are Nash equilibria, as stability is required in addition.

## Perturbations of Equilibria in Evolutionary Game Theory

The evolutionary viewpoint on game theory allows to study Nash equilibria/ESS within the standard framework of dynamical systems theory, by using the concepts of stability, asymptotical stability, global stability and related notions. In fact, one can do more than that: the problem of invasion by a mutant, the biological basis of the ESS concept of Maynard-Smith, can always be formulated in terms of a dynamical coupling of the mutant and the incumbent species and hence studied in terms of the stability of a rest point of a dynamical system. In principle, the same idea can be generalized to simultaneous invasion by more than one mutant and, although the problem may be technically much more difficult, the basic procedure remains the same.

As we did with the fitness landscape concept, when considering perturbations of equilibria, our interest goes beyond this traditional stability ideas, and once again, we need to focus on the deviations from the framework that allows to derive the replicator equation. There are a number of such deviations. The simplest ones are the inclusion of mutations or migrations, leading to the

so-called replicator-mutator equation (Page and Nowak 2002), that can be subsequently studied as a dynamical system. Other deviations affect much more, and in a way more difficult to apprehend, to the evolutionary dynamics and its equilibria, such as considering finite size populations, alternative learning/reproduction dynamics, or the non-universality of interactions among individuals. In this section we will choose this last point as our specific example, and analyze the consequences of relaxing the hypothesis that every player plays every other one. This hypothesis is needed to substitute the payoff earned by a player by what she would have obtained facing the average player of the population (an approach that has been traditionally used in physics under the name of mean-field approximation). However, interactions may not be universal after all, either because of spatial or temporal limitations. We will address both in what follows. The reader is referred to (Nowak 2006) for discussions of other perturbations.

### Spatial Perturbations

One of the reasons why maybe not all individuals interact with all others is that they could not possibly meet. In biological terms this may occur because the population is very sparsely distributed and every individual meets only a few others within its living range, or else in a very numerous population where it is impossible in practice to meet all individuals. In social terms, an alternative view is the existence of a social network or network or contacts that prescribes who interacts with whom.

This idea was first introduced in a famous paper by (Nowak and May 1992) on the evolutionary dynamics of the Prisoner's Dilemma on a square lattice. In the Prisoner's Dilemma two players simultaneously decide cooperate or to defect. Cooperation results in a benefit  $b$  to the recipient but incurs costs  $c$  to the donor ( $b > c > 0$ ). Thus, mutual cooperation pays a net benefit of  $R = b - c$  whereas mutual defection results in  $P = 0$ . However, unilateral defection yields the highest payoff  $T = b$  and the cooperator

has to bear the costs  $S = -c$ . It immediately follows that it is best to defect regardless of the opponents decision. For this reason defection is the evolutionarily stable strategy even though all individuals would be better off if all would cooperate (mutual cooperation is better than mutual defection because  $R > P$ ). Mutual defection is also the only Nash equilibrium of the game. All this translates into the following payoff matrix:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array} \quad (9)$$

For this matrix to correspond to a Prisoner's Dilemma game, the ordering of payoffs must be  $T > R > P > S$ . As we will see below, other orderings define different games.

What Nowak and May did was to set the individuals on the nodes of a square lattice, where they played the game only with their nearest and next-nearest neighbors (Moore neighborhood). They ran simulations with the following dynamics: every individual played the game with her neighbors and collected the corresponding payoff, and afterwards she updated her strategy by imitating that of her most successful (in terms of payoff) neighbor. In their simulations, Nowak and May found that if they started with a population with a majority of cooperators, a large fraction of them remained cooperators instead of changing their behavior towards the ESS, namely, defection. The reason is that the structured interaction allowed cooperators to do well and avoid exploitation by defectors by grouping into clusters, inside which they interacted mostly with other cooperators, whereas defectors at the boundaries of those clusters, interacting mostly with other defectors, did not fare as well and therefore did not induce cooperators to defect.

This result is partly due to the imitation dynamics, which, if postulated to rule the evolution of a population of individuals that interact with all others, does not lead to the replicator equation. As we mentioned in the preceding section, the update rule for the strategies is arbitrary and can be chosen at will (preferably with some specific



modelling in my mind). To reproduce the behavior of the replicator equation, a probabilistic rule has to be used (Gintis 2000), and with this rule the equilibrium is not changed and the population evolves to full defection. However, (Nowak and May 1992) opened the door to a number of more detailed studies that considered also different dynamical rules including the one corresponding to the replicator dynamics in which it was shown that the structure of a population definitely had a strong influence on the game equilibria.

One such study, perhaps the most systematic to date, was carried out by (Hauert 2002), who compared the equilibrium frequencies of cooperators and defectors in populations with and without spatial structuring (square lattices), finding two important results: First, including spatial extension has indeed significant effects on the equilibrium frequencies of cooperators and defectors. In some parameter regions spatial extension promotes cooperative behavior while inhibiting it in others; and, second, differences in the initial frequencies of cooperators are readily leveled out and hardly affect the equilibrium frequencies except for  $T < 1$ ,  $S < 0$ . This choice is not the Prisoner's Dilemma anymore, it corresponds to the so-called Stag Hunt game (Skyrms 2003), and in the replicator dynamics is a bistable system where the initial frequencies determine the long term behavior, a feature that is generally preserved for the spatial setting. Of course, (Hauert 2002) also observed that the size of the neighborhood obviously affects the spreading speed of successful strategies. Interestingly, although the message seems to be that the strategy of cooperation is favored over the strategy of defection by the presence of a spatial structure, this is not always the case, and in games where the equilibrium population has a certain percentage of both types of strategists, the network of interactions makes the frequency of cooperators decrease (Hauert and Doebeli 2004). Therefore, the effect of this perturbation is not all trivial and needs careful consideration.

All the results discussed so far correspond to a square lattice as substrate to define the interaction pattern, but this is certainly a highly idealized setup that can hardly correspond to any real,

natural system. Recent studies have shown that the results also depend on the type of graph or network used. Thus, (Santos et al. 2006) have shown that in more realistic, heterogeneous populations, modeled by random graphs of different types, the sustainability of cooperation (implying the departure of the equilibrium predicted by the replicator equation) is simpler to achieve than in homogeneous populations, a result which is valid irrespective of the dilemma or game adopted as a metaphor of cooperation. Therefore, heterogeneity constitutes a powerful mechanism for the emergence of cooperation (and consequently an important perturbation of the dynamics), since even for mildly heterogeneous populations it leads to sizeable effects in the evolution of cooperation. The overall enhancement of cooperation obtained on single-scale and scale-free graphs (Barabási and Albert 1999) may be understood as resulting from the interplay of two mechanisms: The existence of many long-range connections in random and small-world networks (Watts and Strogatz 1998), which precludes the formation of compact clusters of cooperators, and the heterogeneity exhibited by these networks, which opens a new route for cooperation to emerge and contributes to enhance cooperation (which increases with heterogeneity), counteracting the previous effect. This result depends also on the intricate ties between individuals, even for the same class of graphs, features absent in the replicator dynamics.

We have thus seen that removing the hypothesis of universal interaction is a strong perturbation to equilibrium as understood from the replicator dynamics. There have been many other studies following those we can possibly review here; a recent, very comprehensive summary can be found in (Szabó and Fáth 2007). It must be realized that, intermixed with the effect of the network of interactions, the different dynamical rules one can think of have also a relevant influence on the equilibria. While we have not considered them as perturbations of the replicator dynamics here, because they do not behave as described by that equation even in the presence of universal interactions, it must be kept in mind that they do affect the equilibria and therefore they must be properly



specified in any serious study of evolutionary game theory.

## Time Scales

Let us now come back to the situation in which there is no spatial structure, and every agent can in principle play the game against every other one. Afterwards, reproduction proceeds according to the payoff earned during the game stage. As we have already said, for large populations, this amounts to saying that every player gains the payoff of the game averaged in the current distribution of strategies. In terms of time scales, such an evolution corresponds to a regime in which reproduction-selection events take place at a much slower rate than the interaction between agents. However, these two time scales need not be different in general and, in fact, for many specific applications they can arguably be of the same order (Hendry and Kinnison 1999).

To study different rates of selection we can consider the following new dynamics (Roca et al. 2006, 2007; Sánchez and Cuesta 2005). There is a population with  $N$  players. A pair of individuals is randomly selected for playing, earning each one an amount of fitness according to the rules of the game. This game act is repeated  $s$  times, choosing a new random pair of players in each occasion.

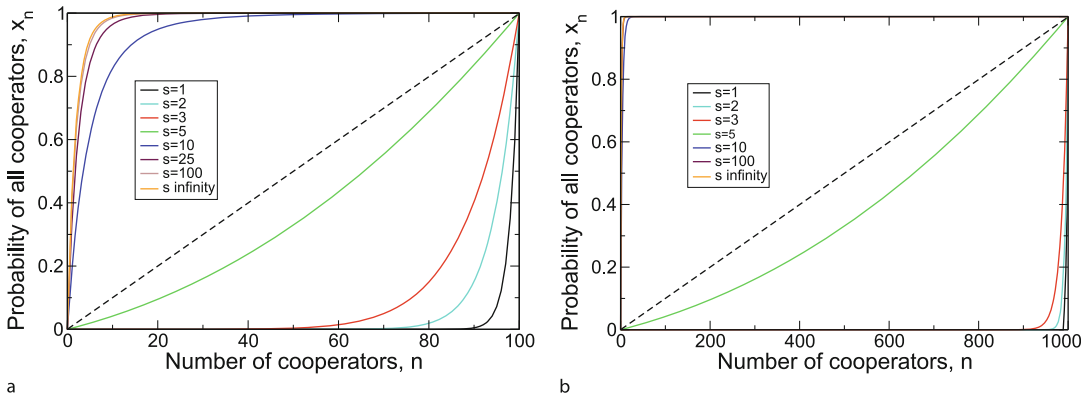
Afterwards, selection takes place. Following (Nowak et al. 2004), we have chosen Moran dynamics (Moran 1962) as the most suitable to model selection in a finite population. This is necessary because the replicator equation is posed for continuous values of the populations and here we need to consider discrete values, i.e., individual by individual, in order to pinpoint the existing time scales. However, it can be shown (Roca et al. 2006) that the equilibria of Moran dynamics are the same as those of the replicator equation, and in fact, the whole evolution is the same except for a rescaling of time. Moran dynamics is defined as follows: One individual among the population of  $N$  players is chosen for reproduction proportionally to its fitness, and its offspring replaces a randomly chosen individual.

As the fitness of all players is set to zero before the following round of  $s$  games, the overall result is that all players have been replaced by one descendant, but the player selected for reproduction has had a reproductive advantage of doubling its offspring at the expense of the randomly selected player. It is worth noting that the population size  $N$  is therefore constant along the evolution.

The parameter  $s$  controls the time scales of the model, i.e. reflects the relation between the rate of selection and the rate of interaction. For  $s \ll N$  selection is very fast and very few individuals interact between reproduction events. Higher values of  $s$  represent proportionally slower rates of selection. Thus, when  $s \gg N$  selection is very slow and population is effectively well-mixed and we recover the behavior predicted by the replicator equation.

The most striking example of the influence of the selection is the so-called Harmony game, a trivial one that has henceforth never been studied, and that is determined by  $R > T > S > P$ . The *only* Nash equilibrium or ESS of this game is mutual cooperation, as it is obvious from the payoffs: The best option for both players is to cooperate, which yields the maximum payoff for each one. Let us denote by  $0 \leq n \leq N$  the number of cooperators present in the population, and look at the probability  $x_n$  of ending up in state  $n = N$  (i.e., all players cooperate) when starting in state  $n < N$ . For  $s = 1$  and  $s \rightarrow \infty$ , an exact, analytical expression for  $x_n$  can be obtained (Roca et al. 2006). For arbitrary values of  $s$ , such a closed form cannot be found; however, it is possible to carry out a combinatorial analysis of the possible combinations of rounds and evaluate, numerically but exactly,  $x_n$ .

In Fig. 2a, we show that the rationally expected outcome of a population consisting entirely of cooperators is not achieved for small and moderate values of  $s$ , our selection rate parameter. For the smallest values, only when starting from a population largely formed by cooperators there is some chance to reach full cooperation; most of the times, defectors will eventually prevail and invade the whole population. This counterintuitive result may arise even for values of  $s$  comparable to the population size, by choosing suitable payoffs. Interestingly, the main result that



**Perturbation of Equilibria in the Mathematical Theory of Evolution, Fig. 2** Probability of ending up with all cooperators starting from  $n$  cooperators,  $x_n$  for different values of  $s$ . (a) For the smallest values of  $s$ , full cooperation is only reached if almost all agents are initially cooperators. Values of  $s$  of the order of 10 show a behavior much more favorable to cooperators. In this plot, the population size is  $N = 100$ . (b) Taking a population of  $N = 1000$ , we observe

that the range of values of  $s$  for which defectors are selected does not depend on the population size, only the shape of the curves does. Parameter choices are: Number of games between reproduction events,  $s$ , as indicated in the plots; payoffs for the Harmony game,  $R = 11$ ,  $S = 2$ ,  $T = 10$ ,  $P = 1$ . The dashed line corresponds to a probability to reach full cooperation equal to the initial fraction of cooperators and is shown for reference

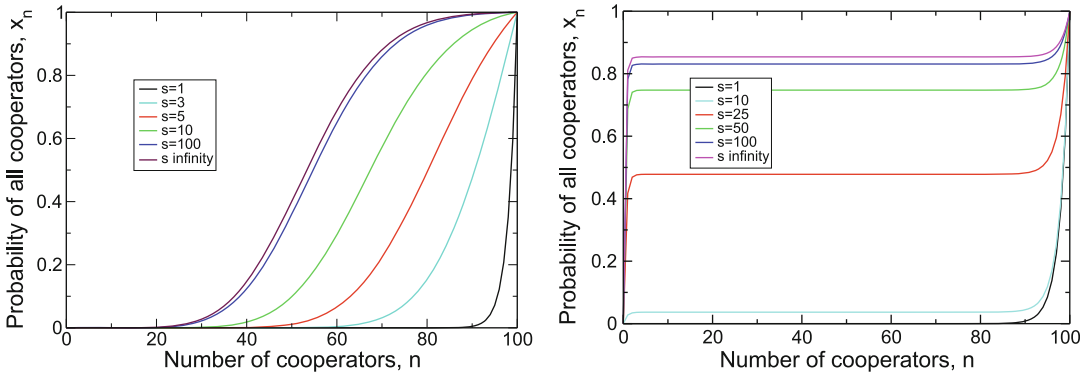
defection is selected for small values of  $s$  does not depend on the population size  $N$ ; only details such as the shape of the curves (cf. Fig. 2b) are modified by  $N$ .

In the preceding paragraph we have chosen the Harmony game to discuss the effect of the rate of selection, but this effect is very general and appears in many other games. Consider the example of the already mentioned Stag-Hunt game (Skyrms 2003), with payoffs  $R > T > P > S$ . This is the paradigmatic situation of game with two Nash equilibria in pure strategies, mutual cooperation and mutual defection, each one with its own basis of attraction in the replicator equation framework (in general, which of these equilibria is selected has been the subject of a long argument in the past, and rationales for both of them can be provided (Skyrms 2003)). As Fig. 3 (left) shows, simulation results for finite  $s$  are largely different from the curve obtained for  $s \rightarrow \infty$ : Indeed, we see that for  $s = 1$ , all agents become defectors except for initial densities close to 1. Even for values of  $s$  as large as  $N$  evolution will more likely lead to a population entirely consisting of defectors.

Yet another example of the importance of the selection rate is provided by the Snowdrift game, with payoffs  $T > R > S > P$ . Figure 3 (right)

shows that for small values of  $s$  defectors are selected for almost any initial fraction of cooperators. When  $s$  increases, we observe an intermediate regime where both full cooperation and full defection have nonzero probability, which, interestingly, is almost independent of the initial population. And, for large enough  $s$ , full cooperation is almost always achieved.

With these examples, it is clear that considering independent interaction and selection time scales may lead to highly non-trivial, counter-intuitive results (Roca et al. 2006). showed that out of the 12 possible different games with two players, six were severely changed by the introduction of the game time scale, whereas the other six remained with the same equilibrium structure. Of course, the extent of the modifications of the replicator dynamics picture depends on the structure of the unperturbed phase space. Thus, rapid selection perturbations show up in changes of the asymptotically selected equilibria, i.e., of the asymptotically stable one, to changes of the basins of attraction of equilibria, or to suppression of long-lived metastable equilibria. As in the case of spatial perturbations, we are thus faced with a most relevant influence on the equilibria of the evolutionary game.



**Perturbation of Equilibria in the Mathematical Theory of Evolution, Fig. 3** Left: Same as Fig. 2 for the Stag-Hunt game. The probability of ending up with all cooperators starting from  $n$  cooperators,  $x_n$ , is very low when  $s$  is small, and as  $s$  increases it tends to a quasi-symmetric distribution around  $1/2$ . Payoffs for the Stag-Hunt game,  $R = 6$ ,  $S = 1$ ,  $T = 5$ ,  $P = 2$ . Right: Same as Fig. 2 for the Snowdrift game. The probability of ending up

with all cooperators starting from  $n$  cooperators is almost independent of  $n$  except for very small or very large values. Small  $s$  values lead once again to selection of defectors, whereas cooperators prevail more often as  $s$  increases. Payoffs for the Snowdrift game,  $R = 1$ ,  $S = 0.35$ ,  $T = 1.65$ ,  $P = 0$ . Other parameter choices are: Population,  $N = 100$ ; number of games between reproduction events,  $s$ , as indicated in the plot

## Future Directions

As we have seen in this necessarily short excursion, the simplest mathematical models of evolution allow for a detailed, analytical study of their equilibria (which are supposed to represent stable states of populations) but, when leaving aside some of the hypothesis involved in the derivation of those simple models, the structure of equilibria may be seriously modified and highly counter-intuitive results may arise. We have not attempted to cover all possible perturbations but we believe we have provided evidence enough that their effect is certainly very relevant. When trying to bridge the gap between simple models and reality, other hypotheses will be broken, maybe more than one simultaneously, and subsequently the equilibria will be severely affected.

In the future, we believe that this line of research will undergo very interesting developments, particularly in the framework of evolutionary game theory, as the fitness landscape picture seems to be rather well understood and, on the other hand, is felt to be a much too simple model. In the case of evolutionary games, while some of the results we have collected here are analytical, there are many others which are only numerical, in particular when perturbations depending on space (evolutionary

game theory on graphs) are considered. A lot of research needs to be devoted in the next few years to understand this problem analytically, more so because there is now a “zoo” of results that are even hard to classify or interpret within a common basis. This will probably require a combined effort from different mathematical disciplines, ranging from discrete mathematics to dynamical systems through, of course, graph theory.

On the other hand, our examples have consisted of two-player, two-strategy, symmetric games, i.e., the simplest possible scenario. There are practically no results about games with more than two strategies or, even worse, with more than two players. In fact, even the classification of the phase portraits within the replicator equations for those situations is far from understood, the more so the higher the dimensionality of the problem. Much remains to be done in this direction. Asymmetrical games are a different story; for those, the replicator equation is not (6) anymore but rather one has to take into account that payoffs when playing as player 1 or player 2 are not the same, and the corresponding equation is more complicated. Again, this line of research is still in its infancy and awaiting for dedicated work.

Finally, a very interesting direction is the application of the results to problems in social or

biological contexts. The evolutionary game theory community has been relying strongly on the predictions from the replicator equation which we now see may not agree with reality or at least with what occurs when some of its hypotheses are not fulfilled. This has led to a number of conundrums, particularly prominent among those being the problem of the emergence of cooperation. A recent study (Sánchez and Cuesta 2005) have shown that, in a scenario described by the so-called Ultimatum game, taking into consideration the possible separation of time scales leads to results compatible with the experimental observations on human subjects, observations that the replicator equation is not able to reproduce. We envisage that similar results will be ubiquitous when trying to match the predictions of the replicator equation with actual systems or problems. Understanding the effect of perturbations in a comprehensive manner will then be the key to the fruitful development of the theory as a “natural” or “physical” one.

## Bibliography

### Primary Literature

- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
- Drossel B (2001) Biological evolution and statistical physics. *Adv Phys* 50:209–295
- Fisher RA (1958) The genetical theory of natural selection, 2nd edn. Dover, New York
- Gintis H (2000) Game theory evolving. Princeton University, Princeton
- Hauert C (2002) Effects of space in 2 x 2 games. *Int J Bifurc Chaos* 12:1531–1548
- Hauert C, Doebeli M (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428:643–646
- Hendry AP, Kinnison MT (1999) The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53:1637–1653
- Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University, Cambridge

- Hofbauer J, Sigmund K (2003) Evolutionary game dynamics. *Bull Am Math Soc* 40:479–519
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University, Cambridge
- Maynard-Smith J (1982) Evolution and the theory of games. Cambridge University, Cambridge
- Maynard-Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246:15–18
- Moran PAP (1962) The statistical processes of evolutionary theory. Clarendon, Oxford
- Nash JF (1950) Equilibrium points in n-person games. *Proc Natl Acad Sci U S A* 36:48–49
- Nowak MA (2006) Evolutionary dynamics: exploring the equations of life. Harvard University Press, Harvard
- Nowak MA, May R (1992) Evolutionary games and spatial chaos. *Nature* 359:826–829
- Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646–650
- Page K, Nowak MA (2002) A unified evolutionary dynamics. *J Theor Biol* 219:93–98
- Roca CP, Cuesta JA, Sánchez A (2006) Time scales in evolutionary dynamics. *Phys Rev Lett* 97: art. no. 158701
- Roca CP, Cuesta JA, Sánchez A (2007) The importance of selection rate in the evolution of cooperation. *Eur Phys J Special Topics* 143:51–58
- Sánchez A, Cuesta JA (2005) Altruism may arise by individual selection. *J Theor Biol* 235:233–240
- Santos FC, Pacheco JM, Lenaerts T (2006) Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc Natl Acad Sci U S A* 103: 3490–3494
- Skyrms B (2003) The stag hunt and the evolution of social structure. Cambridge University, Cambridge
- Szabó G, Fáth G (2007) Evolutionary games on graphs. *Phys Rep* 446:97–216
- Watts DJ, Strogatz SH (1998) Collective dynamics of “Small World” networks. *Nature* 393:440–442
- Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc 6th Int Cong Genet* 1:356–366

### Books and Reviews

- Nowak M, Sigmund K (2004) Evolutionary dynamics of biological games. *Science* 303:793–799
- Taylor PD, Jonker L (1978) Evolutionarily stable strategies and game dynamics. *J Math Biosci* 40:145–156
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton

# Perturbation Theory for Non-smooth Systems

Marco Antônio Teixeira  
Department of Mathematics, Universidade  
Estadual de Campinas, Campinas, Brazil

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
Preliminaries  
Discontinuous Systems  
Singular Perturbation Problem  
Regularization Process  
Vector Fields Near the Boundary  
A Construction  
Codimension-one M-Singularity in Dimensions  
Two and Three  
Singular Perturbation Problem in 2D  
Future Directions  
Some Problems  
Conclusion  
Bibliography

## Glossary

**Non-smooth dynamical system** Systems derived from ordinary differential equations when the non-uniqueness of solutions is allowed. In this article we deal with discontinuous vector fields in  $R^n$  where the discontinuities are concentrated in a codimension-one surface.

**Bifurcation** In a  $k$ -parameter family of systems, a bifurcation is a parameter value at which the phase portrait is not structurally stable.

**Typical singularity** Are points on the discontinuity set where the orbits of the system through them must be distinguished.

## Definition of the Subject

In this article we survey some qualitative and geometric aspects of non-smooth dynamical systems theory. Our goal is to provide an overview of the state of the art on the theory of contact between a vector field and a manifold, and on discontinuous vector fields and their perturbations. We also establish a bridge between two-dimensional non-smooth systems and the geometric singular perturbation theory. Non-smooth dynamical systems is a subject that has been developing at a very fast pace in recent years due to various factors: its mathematical beauty, its strong relationship with other branches of science and the challenge in establishing reasonable and consistent definitions and conventions. It has become certainly one of the common frontiers between mathematics and physics/engineering. We mention that certain phenomena in control systems, impact in mechanical systems and nonlinear oscillations are the main sources of motivation for our study concerning the dynamics of those systems that emerge from differential equations with discontinuous right-hand sides. We understand that non-smooth systems are driven by applications and they play an intrinsic role in a wide range of technological areas.

## Introduction

The purpose of this article is to present some aspects of the geometric theory of a class of non-smooth systems. Our main concern is to bring the theory into the domain of geometry and topology in a comprehensive mathematical manner.

Since this is an impossible task, we do not attempt to touch upon all sides of this subject in one article. We focus on exploring the local behavior of systems around typical singularities. The first task is to describe a generic persistence

of a local theory (structural stability and bifurcation) for discontinuous systems mainly in the two- and three-dimensional cases. Afterwards we present some striking features and results of the regularization process of two-dimensional discontinuous systems in the framework developed by Sotomayor and Teixeira (Sotomayor and Teixeira 1996) and establish a bridge between those systems and the fundamental role played by the Geometric Singular Perturbation Theory (GSPT). This transition was introduced in (Buzzi et al. 2006) and we reproduce here its main features in the two-dimensional case. For an introductory reading on the methods of geometric singular perturbation theory we refer to (Dumortier and Roussarie 1996; Fenichel 1979; Jones 1995). In section “Definition of the Subject” we introduce the setting of this article. In section “Introduction” we survey the state of the art of the contact between a vector field and a manifold. The results contained in this section are crucial for the development of our approach. In section “Preliminaries” we discuss the classification of typical singularities of non-smooth vector fields. The study of non-smooth systems, via GSPT, is presented in section “Vector Fields near the Boundary”. In section “Generic Bifurcation” some theoretical open problems are presented.

One aspect of the qualitative point of view is the problem of structural stability, the most comprehensive of many different notions of stability. This theme was studied in 1937 by Andronov–Pontryagin (see Andronov and Pontryagin 1937). This problem is of obvious importance, since in practice one obtains a lot of qualitative information not only on a fixed system but also on its nearby systems.

We deal with non-smooth vector fields in  $R^{n+1}$  having a codimension-one submanifold  $M$  as its discontinuity set. The scheme in this work toward a systematic classification of typical singularities of non-smooth systems follows the ideas developed by Sotomayor–Teixeira in (Sotomayor and Teixeira 1988) where the problem of contact between a vector field and the boundary of a manifold was discussed. Our approach intends to be self-contained and is accompanied by an extensive bibliography. We will try to focus here on

areas that are complementary to some recent reviews made elsewhere.

The concept of structural stability in the space of non-smooth vector fields is based on the following definition:

**Definition 1** Two vector fields  $Z$  and  $\tilde{Z}$  are  $C^0$  equivalent if there is an  $M$ -invariant homeomorphism  $h : R^{n+1} \rightarrow R^{n+1}$  that sends orbits of  $Z$  to orbits of  $\tilde{Z}$ .

A general discussion is presented to study certain unstable non-smooth vector fields within a generic context. The framework in which we shall pursue these unstable systems is sometimes called generic bifurcation theory. In (Andronov and Pontryagin 1937) the concept of  $k$ th-order structural stability is also presented; in a local approach such setting gives rise to the notion of a codimension- $k$  singularity. In studies of classical dynamical systems, normal form theory has been well accepted as a powerful tool in studying the local theory (see Arnold 1983). Observe that, so far, bifurcation and normal form theories for non-smooth vector fields have not been extensively studied in a systematic way.

Control Theory is a natural source of mathematical models of these systems (see, for instance, Andronov et al. 1966; Bonnard and Chyba 2000; Flügge-Lotz 1953; Seidman 2006; Sussmann 1979). Interesting problems concerning discontinuous systems can be formulated in systems with hysteresis (Seidman 2006), economics (Henry 1972; Ito 1979) and biology (Bazykin 1998). It is worth mentioning that in (Anosov 1959) a class of relay systems in  $R^n$  is discussed. They have the form:

$$\dot{X} = Ax + \operatorname{sgn}(x_1)k$$

where  $x = (x_1, x_2, \dots, x_n)$ ,  $A \in M_R(n, n)$  and  $k = (k_1, k_2, \dots, k_n)$  is a constant vector in  $R^n$ . In (Jacquemard and Teixeira 2003b, 2005) the generic singularities of reversible relay systems in  $4D$  were classified. In (Zelikin and Borisov 1994) some properties of non-smooth dynamics are discussed in order to understand some



phenomena that arise in chattering control. We mention the presence of chaotic behavior in some non-smooth systems (see for example Chua 1992). It is worthwhile to cite (Ekeland 1977), where the main problem in the classical calculus of variations was carried out to study discontinuous Hamiltonian vector fields. We refer to (di Bernardo et al. 2005) for a comprehensive text involving non-smooth systems which includes many models and applications. In particular motivating models of several non-smooth dynamical systems arising in the occurrence of impacting motion in mechanical systems, switchings in electronic systems and hybrid dynamics in control systems are presented together with an extensive literature on impact oscillators which we do not attempt to survey here. For further reading on some mathematical aspects of this subject we recommend (Chillingworth 2002) and references therein. A setting of general aspects of non-smooth systems can be found also in (Luo 2006) and references therein. Our discussion does not focus on continuous but rather on non-smooth dynamical systems and we are aware that the interest in this subject goes beyond the approach adopted here.

The author wishes to thank R. Garcia, T.M. Seara and J. Sotomayor for many helpful conversations.

## Preliminaries

Now we introduce some of the terminology, basic concepts and some results that will be used in the sequel.

**Definition 2** Two vector fields  $Z$  and  $\tilde{Z}$  on  $R^n$  with  $Z(0) = \tilde{Z}(0)$  are germ-equivalent if they coincide on some neighborhood  $V$  of 0.

The equivalent classes for this equivalence are called germs of vector fields. In the same way as defined above, we may define germs of functions. For simplicity we are considering the germ notation and we will not distinguish a germ of a function and any one of its representatives. So, for example, the notation  $h : R^n, 0 \rightarrow R$  means

that the  $h$  is a germ of a function defined in a neighborhood of 0 in  $R^n$ . Referto (Dumortier 1977) for a brief and nice introduction of the concepts of *germ* and *k-jet* of functions.

## Discontinuous Systems

Let  $M = h^{-1}(0)$ , where  $h$  is (a germ of) a smooth function  $h : R^{n+1}, 0 \rightarrow R$  having  $0 \in R$  as its regular value. We assume that  $0 \in M$ .

Designate by  $\chi(n+1)$  the space of all germs of  $C^r$  vector fields on  $R^{n+1}$  at 0 endowed with the  $C^r$ -topology with  $r > 1$  and large enough for our purposes. Call  $\Omega(n+1)$  the space of all germs of vector fields  $Z$  in  $R^{n+1}, 0$  such that

$$Z(q) = \begin{cases} X(q), & \text{for } h(q) > 0, \\ Y(q), & \text{for } h(q) < 0, \end{cases} \quad (1)$$

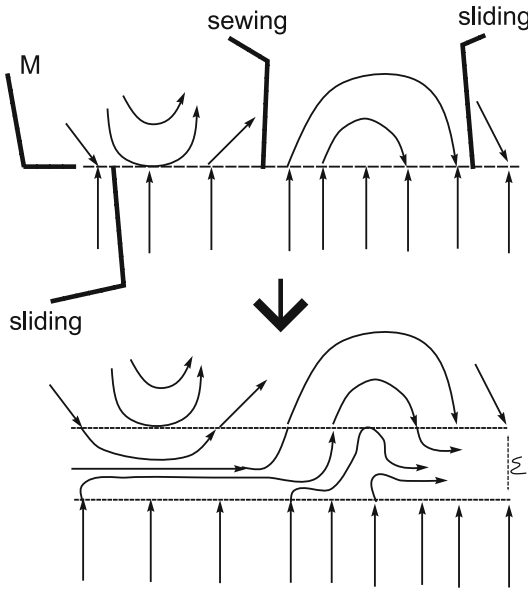
The above field is denoted by  $Z = (X, Y)$ . So we are considering  $\Omega(n+1) = \chi(n+1) \times \chi(n+1)$  endowed with the product topology.

**Definition 3** We say that  $Z \in \Omega(n+1)$  is structurally stable if there exists a neighborhood  $U$  of  $Z$  in  $\Omega(n+1)$  such that every  $\tilde{Z} \in U$  is  $C^0$ -equivalent with  $Z$ .

To define the orbit solutions of  $Z$  on the switching surface  $M$  we take a pragmatic approach. In a well characterized open set  $O$  of  $M$  (described below) the solution of  $Z$  through a point  $p \in O$  obeys the Filippov rules and on  $M - O$  we accept it to be multivalued. Roughly speaking, as we are interested in studying the structural stability in  $\Omega(n+1)$  it is convenient to take into account all the leaves of the foliation in  $R^{n+1}$  generated by the orbits of  $Z$  (and also the orbits of  $X$  and  $Y$ ) passing through  $p \in M$ . (see Fig. 1).

The trajectories of  $Z$  are the solutions of the autonomous differential system  $\dot{q} = Z(q)$ .

In what follows we illustrate our terminology by presenting a simplified model that is found in the classical electromagnetism theory (see for instance Jackson 1999):



**Perturbation Theory for Non-smooth Systems, Fig. 1** A discontinuous system and its regularization

$$\ddot{x} - \dots x + \alpha \operatorname{sign} x = 0.$$

with  $\alpha > 0$ .

So this system can be expressed by the following objects:  $h(x, y, z) = x$  and  $Z = (X, Y)$  with  $X(x, y, z) = (y, z, z + \alpha)$  and  $Y(x, y, z) = (y, z, z - \alpha)$ .

For each  $X \in \chi(n+1)$  we define the smooth function  $Xh : R^{n+1} \rightarrow R$  given by  $Xh = X \cdot \nabla h$  where  $\cdot$  is the canonical scalar product in  $R^{n+1}$ .

We distinguish the following regions on the discontinuity set  $M$ :

- (i)  $M_1$  is the *sewing region* that is represented by  $h = 0$  and  $(Xh)(Yh) > 0$ ;
- (ii)  $M_2$  is the *escaping region* that is represented by  $h = 0$ ,  $(Xh) > 0$  and  $(Yh) < 0$ ;
- (iii)  $M_3$  is the *sliding region* that is represented by  $h = 0$ ,  $(Xh) < 0$  and  $(Yh) > 0$ .

We set  $O = \cup_{i=1,2,3} M_i$ .

Consider  $Z = (X, Y) \in \Omega(n+1)$  and  $p \in M_3$ . In this case, following Filippov's convention, the solution  $\gamma(t)$  of  $Z$  through  $p$  follows, for  $t \geq 0$ , the orbit of a vector field tangent to  $M$ . Such system is called *sliding vector field* associated with  $Z$  and it will be defined below.

**Definition 4** The sliding vector field associated to  $Z = (X, Y)$  is the smooth vector field  $Z^s$  tangent to  $M$  and defined at  $q \in M_3$  by  $Z^s(q) = m - q$  with  $m$  being the point where the segment joining  $q + X(q)$  and  $q + Y(q)$  is tangent to  $M$ .

It is clear that if  $q \in M_3$  then  $q \in M_2$  for  $-Z$  and then we define the escaping vector field on  $M_2$  associated with  $Z$  by  $Z^e = -(-Z)^s$ . In what follows we use the notation  $Z^M$  for both cases.

We recall that sometimes  $Z^M$  is defined in an open region  $U$  with boundary. In this case it can be  $C^r$  extended to a full neighborhood of  $p \in \partial U$  in  $M$ .

When the vectors  $X(p)$  and  $Y(p)$ , with  $p \in M_2 \cup M_3$  are linearly dependent then  $Z^M(p) = 0$ . In this case we say that  $p$  is a simple singularity of  $Z$ . The other singularities of  $Z$  are concentrated outside the set  $O$ .

We finish this subsection with a three-dimensional example:

Let  $Z = (X, Y) \in \Omega(3)$  with  $h(x, y, z) = z$ ,  $X = (1, 0, x)$  and  $Y = (0, 1, y)$ . The system determines four quadrants around 0, bounded by  $\tau_X = \{x = 0\}$  and  $\tau_Y = \{y = 0\}$ . They are:  $Q_1^+ = \{x > 0, y > 0\}$ ,  $Q_1^- = \{x < 0, y < 0\}$ ,  $Q_2 = \{x < 0, y > 0\}$  (sliding region) and  $Q_3 = \{x > 0, y < 0\}$  (escaping region). Observe that  $M_1 = Q_1^+ \cup Q_1^-$ .

The sliding vector field defined in  $Q_2$  is expressed by:

$$Z^s(x, y, z) = (y - x)^{-1} \left( x + y, \frac{y + x}{8}, 0 \right).$$

Such a system is (in  $Q_2$ ) equivalent to  $G(x, y, z) = \left( x + y, \frac{y + x}{8}, 0 \right)$ . In our terminology we consider  $G$  a smooth extension of  $Z^s$ , that is defined in a whole neighborhood of 0. It is worthwhile to say that  $G$  is in fact a system which is equivalent to the original system in  $Q_2$ .

In (Teixeira 1993) a generic classification of one-parameter families of sliding vector fields is presented.

## Singular Perturbation Problem

A singular perturbation problem is expressed by a differential equation  $z' = \alpha(z, \varepsilon)$  (refer to

Dumortier and Roussarie 1996; Fenichel 1979; Jones 1995) where  $z \in R^{n+m}$ ,  $\varepsilon$  is a small non-negative real number and  $\alpha$  is a  $C^\infty$  mapping.

Let  $z = (x, y) \in R^{n+m}$  and  $f: R^{m+n} \rightarrow R^m$ ,  $g: R^{m+n} \rightarrow R^n$  be smooth mappings. We deal with equations that may be written in the form

$$\begin{cases} x' = f(x, y, \varepsilon) \\ y' = \varepsilon g(x, y, \varepsilon) \end{cases} \quad x = x(\tau), y = y(\tau). \quad (2)$$

An interesting model of such systems can be obtained from the singular van der Pol's equation

$$\varepsilon x'' + (x^2 + x)x' + x - a = 0. \quad (3)$$

The main trick in the geometric singular perturbation (GSP) is to consider the family (2) in addition to the family

$$\begin{cases} \varepsilon \dot{x} = f(x, y, \varepsilon) \\ \dot{y} = g(x, y, \varepsilon) \end{cases} \quad x = x(t), y = y(t) \quad (4)$$

obtained after the time rescaling  $t = \varepsilon \tau$ .

Equation (2) is called the fast system and (4) the slow system. Observe that for  $\varepsilon > 0$  the phase portrait of fast and slow systems coincide.

For  $\varepsilon = 0$ , let  $S$  be the set of all singular points of (2). We call  $S$  the slow manifold of the singular perturbation problem and it is important to notice that Eq. (4) defines a dynamical system on  $S$  called the *reduced problem*.

Combining results on the dynamics of these two limiting problems (2) and (4), with  $\varepsilon = 0$ , one obtains information on the dynamics for small values of  $\varepsilon$ . In fact, such techniques can be exploited to formally construct approximated solutions on pieces of curves that satisfy some limiting version of the original equation as  $\varepsilon$  goes to zero.

**Definition 5** Let  $A, B \subset R^{n+m}$  be compact sets. The Hausdorff distance between  $A$  and  $B$  is  $D(A, B) = \max_{z_1 \in A, z_2 \in B} \{d(z_1, B), d(z_2, A)\}$ .

The main question in GSP-theory is to exhibit conditions under which a singular orbit can be approximated by regular orbits for  $\varepsilon \downarrow 0$ , with respect to the Hausdorff distance.

## Regularization Process

An approximation of the discontinuous vector field  $Z = (X, Y)$  by a one-parameter family of continuous vector fields will be called a regularization of  $Z$ . In (Sotomayor and Teixeira 1996), Sotomayor and Teixeira introduced the regularization procedure of a discontinuous vector field. A transition function is used to average  $X$  and  $Y$  in order to get a family of continuous vector fields that approximates the discontinuous one. Figure 1 gives a clear illustration of the regularization process.

$$\text{Let } Z = (X, Y) \in \Omega(n+1).$$

**Definition 6** A  $C^\infty$  function  $\varphi: R \rightarrow R$  is a transition function if  $\varphi(x) = -1$  for  $x \leq -1$ ,  $\varphi(x) = 1$  for  $x \geq 1$  and  $\varphi'(x) > 0$  if  $x \in (-1, 1)$ . The  $\phi$ -regularization of  $Z = (X, Y)$  is the one-parameter family  $X_\varepsilon \in C^r$  given by

$$\begin{aligned} Z_\varepsilon(q) = & \left( \frac{1}{2} + \frac{\varphi_\varepsilon(h(q))}{2} \right) X(q) \\ & + \left( \frac{1}{2} - \frac{\varphi_\varepsilon(h(q))}{2} \right) Y(q). \end{aligned} \quad (5)$$

with  $h$  given in the above subsection “Discontinuous Systems” and  $\varphi_\varepsilon(x) = \varphi(x/\varepsilon)$ , for  $\varepsilon > 0$ .

As already said before, a point in the phase space which moves on an orbit of  $Z$  crosses  $M$  when it reaches the region  $M_1$ . Solutions of  $Z$  through points of  $M_3$ , will remain in  $M$  in forward time. Analogously, solutions of  $Z$  through points of  $M_2$  will remain in  $M$  in backward time. In (Llibre and Teixeira 1997; Sotomayor and Teixeira 1996) such conventions are justified by the regularization method in dimensions two and three respectively.

## Vector Fields Near the Boundary

In this section we discuss the behavior of smooth vector fields in  $R^{n+1}$  relative to a codimension-one submanifold (say, the above defined  $M$ ). We base our approach on the concepts and results

contained in (Sotomayor and Teixeira 1988; Vishik 1972). The principal advantage of this setting is that the generic contact between a smooth vector field and  $M$  can often be easily recognized. As an application the typical singularities of a discontinuous system can be further classified in a straightforward way.

We say that  $X, Y \in \chi(n+1)$  are  $M$ -equivalent if there exists an  $M$ -preserving homeomorphism  $h: R^{n+1}, 0 \rightarrow R^{n+1}, 0$  that sends orbits of  $X$  into orbits of  $Y$ . In this way we get the concept of  $M$ -structural stability in  $\chi(n+1)$ .

We call  $\Gamma_0(n+1)$  the set of elements  $X$  in  $\chi(n+1)$  satisfying one of the following conditions:

0)  $Xh(0) \neq 0$  (0 is a regular point of  $X$ ). In this case  $X$  is transversal to  $M$  at 0.

1)  $Xh(0) = 0$  and  $X^2h(0) \neq 0$  (0 is a 2-fold point of  $X$ );

2)  $Xh(0) = X^2h(0) = 0, X^3h(0) \neq 0$  and the set  $\{Dh(0), DXh(0), DX^2h(0)\}$  is linearly independent (0 is a cusp point of  $X$ );

...

3)  $Xh(0) = X^2h(0) = \dots = X^n h(0) = 0$  and  $X^{n+1}h(0) \neq 0$ . Moreover the set  $\{Dh(0), DXh(0), \dots, DX^n h(0)\}$  is linearly independent, and 0 is a regular point of the mapping  $Xh|_M$ .

We say that 0 is an  $M$ -singularity of  $X$  if  $h(0) = Xh(0) = 0$ . It is a *codimension-zero*  $M$ -singularity provided that  $X \in \Gamma_0(n+1)$ .

We know that  $\Gamma_0(n+1)$  is an open and dense set in  $\chi(n+1)$  and it coincides with the  $M$ -structurally stable vector fields in  $\chi(n+1)$  (see Vishik 1972).

Denote by  $\tau_X \subset M$  the  $M$ -singular set of  $X \in \chi(n+1)$ ; this set is represented by the equations  $h = Xh = 0$ . It is worthwhile to point out that, generically, all two-folds constitute an open and dense subset of  $\tau_X$ . Observe that if  $X(0) = 0$  then  $X \notin \Gamma_0(n+1)$ .

The  $M$ -bifurcation set is represented by  $\chi_1(n+1) = \chi(n+1) - \Gamma_0(n+1)$ .

Vishik in (Vishik 1972) exhibited the normal forms of a *codimension-zero*  $M$ -singularity. They are:

I) Straightened vector field

$$X = (1, 0, \dots, 0)$$

and

$$\begin{aligned} h(x) &= x_1^{k+1} + x_2 x_1^{k-1} + x_3 x_1^{k-2} \\ &\quad + \dots + x_{k+1}, \\ k &= 0, 1, \dots, n \end{aligned}$$

or

II) Straightened boundary

$$h(x) = x_1$$

and

$$X(x) = (x_2, x_3, \dots, x_k, 1, 0, 0, \dots, 0)$$

We now discuss an important interaction between vector fields near  $M$  and singularities of mapping theory. We discuss how singularity-theoretic techniques help the understanding of the dynamics of our systems.

We outline this setting, which will be very useful in the sequel. The starting point is the following construction.

## A Construction

Let  $X \in \chi(n+1)$ . Consider a coordinate system  $x = (x_1, x_2, \dots, x_{n+1})$  in  $R^{n+1}, 0$  such that

$$M = \{x_1 = 0\}$$

and

$$X = (X^1, X^2, \dots, X^{n+1})$$

Assume that  $X(0) \neq 0$  and  $X^1(0) = 0$ . Let  $N_0$  be any transversal section to  $X$  at 0.

By the implicit function theorem, we derive that:

for each  $p \in M, 0$  there exists a unique  $t = t(p)$  in  $R, 0$  such that the orbit-solution  $t \mapsto \gamma(p, t)$  of  $X$  through  $p$  meets  $N_0$  at a point  $\tilde{p} = \gamma(p, t(p))$ .

We define the smooth mapping  $\rho_X: R^n, 0 \rightarrow R^n$ , by  $\rho_X(p) = \tilde{p}$ . This mapping is a powerful tool in the study of vector fields around the boundary of a manifold (refer to Garcia and Teixeira 2004;

Sotomayor 1974; Sotomayor and Teixeira 1988; Teixeira 1977; Vishik 1972). We observe that  $\tau_X$  coincides with the singular set of  $\rho_X$ .

The late construction implements the following method. If we are interested in finding an equivalence between two vector fields which preserve  $M$ , then the problem can be sometimes reduced to finding an equivalence between  $\rho_X$  and  $\rho_Y$  in the sense of singularities of mappings.

We recall that when 0 is a fold  $M$ -singularity of  $X$  then associated to the fold mapping  $\rho_X$  there is the symmetric diffeomorphism  $\beta_X$  that satisfies  $\rho_X \circ \beta_X = \rho_X$ .

Given  $Z = (X, Y) \in \Omega(n+1)$  such that  $\rho_X$  and  $\rho_Y$  are fold mappings with  $X^2h(0) < 0$  and  $Y^2h(0) > 0$  then the composition of the associated symmetric mappings  $\beta_X$  and  $\beta_Y$  provides a first return mapping  $\beta_Z$  associated to  $Z$  and  $M$ . This situation is usually called a *distinguished fold-singularity*, and the mapping  $\beta_Z$  plays a fundamental role in the study of the dynamics of  $Z$ .

## Codimension-one $M$ -Singularity in Dimensions Two and Three

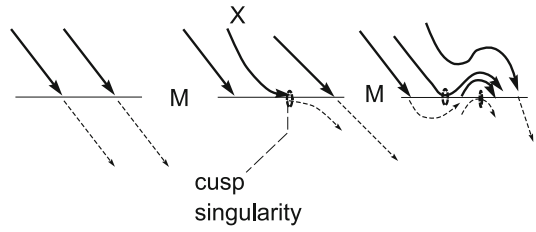
### Case $n = 1$

In this case the unique codimension-zero  $M$ -singularity is a fold point in  $R^2, 0$ . The codimension-one  $M$ -singularities are represented by the subset  $\Gamma_1(2)$  of  $\chi_1(2)$  and it is defined as follows.

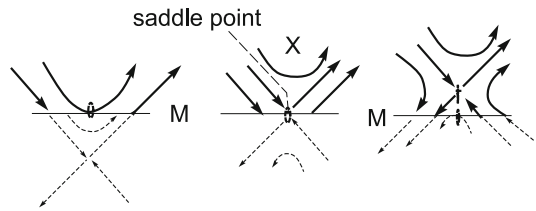
**Definition 7** A codimension-one  $M$ -singularity of  $X \in \Gamma_1(2)$  is either a *cuspl* singularity or an  $M$ -hyperbolic critical point  $p$  in  $M$  of the vector field  $X$ . A cuspl singularity (illustrated in Fig. 2) is characterized by  $Xh(p) = X^2h(p) = 0$ ,  $X^3h(p) \neq 0$ . In the second case this means that  $p$  is a hyperbolic critical point (illustrated in Fig. 3) of  $X$  with distinct eigenvalues and with invariant manifolds (stable, unstable and strong stable and strong unstable) transversal to  $M$ .

In this subsection we consider a coordinate system in  $R^2, 0$  such that  $h(x, y) = y$ .

The next result was proved in (Teixeira 1977). It presents the normal forms of the codimension-one singularities defined above.



**Perturbation Theory for Non-smooth Systems, Fig. 2** The cuspl singularity and its unfolding



**Perturbation Theory for Non-smooth Systems, Fig. 3** The saddle point in the boundary and its unfolding

**Theorem 8** Let  $X \in \chi_1(2)$ . The vector field  $X$  is  $M$ -structurally stable relative to  $\chi_1(2)$  if and only if  $X \in \Gamma_1(2)$ . Moreover,  $\Gamma_1(2)$  is an embedded codimension-one sub manifold and dense in  $\chi_1(2)$ . We still require that any one-parameter family  $X_\lambda$ , ( $\lambda \in (-\varepsilon, \varepsilon)$ ) in  $\chi(1)$  transverse to  $\Gamma_1(2)$  at  $X_0$ , has one of the following normal forms:

- 0.1:  $X_\lambda(x, y) = (1, 0)$  (regular point);
- 0.2:  $X_\lambda(x, y) = (1, x)$  (fold singularity);
- 1.1:  $X_\lambda(x, y) = (1, \lambda + x^2)$  (cuspl singularity);
- 1.2:  $X_\lambda(x, y) = (ax, x + by + \lambda)$ ,  $a = \pm 1$ ,  $b = \pm 2$ ;
- 1.3:  $X_\lambda(x, y) = (x, x - y + \lambda)$ ;
- 1.4:  $X_\lambda(x, y) = (x + y, -x + y + \lambda)$ .

### Case $n = 2$

**Definition 9** A vector field  $X \in \chi(3)$  belongs to the set  $\Gamma_1(a)$  if the following conditions hold:

- (i)  $X(0) = 0$  and 0 is a hyperbolic critical point of  $X$ ;
- (ii) the eigenvalues of  $DX(0)$  are pairwise distinct and the corresponding eigenspaces are transversal to  $M$  at 0,;
- (iii) each pair of non complex conjugate eigenvalues of  $DX(0)$  has distinct real parts.

**Definition 10** A vector field  $X \in \chi(3)$  belongs to the set  $\Gamma_1(b)$  if  $X(0) \neq 0$ ,  $Xh(0) = 0$ ,  $X^2h(0) = 0$  and one of the following conditions hold:

- (1)  $X^3h(0) \neq 0$ ,  $\text{rank} \{Dh(0), DXh(0), DX^2h(0)\} = 2$  and 0 is a non-degenerate critical point of  $Xh|_M$ .
- (2)  $X^3h(0) = 0$ ,  $X^4h(0) \neq 0$  and 0 is a regular point of  $Xh|_M$ .

The next results can be found in (Sotomayor and Teixeira 1988).

**Theorem 11** The following statements hold:

- (i)  $\Gamma_1(3) = \Gamma_1(a) \cup \Gamma_2(b)$  is a codimension-one submanifold of  $\chi(3)$ .
- (ii)  $\Gamma_1(3)$  is open and dense set in  $\chi_1(3)$  in the topology induced from  $\chi_1(3)$ .
- (iii) For a residual set of smooth curves  $\gamma : R, 0 \rightarrow \chi(3)$ ,  $\gamma$  meets  $\Gamma_1(3)$  transversally.

Throughout this subsection we fix the function  $h(x, y, z) = z$ .

**Lemma 12** (Classification Lemma)

The elements of  $\Gamma_1(3)$  are classified as follows:

|                                                    |                                                                                                                                                                                                                           |
|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>(a<sub>11</sub>)</b> Nodal M-Singularity:       | $X(0) = 0$ , the eigenvalues of $DX(0)$ , $\lambda_1, \lambda_2$ , and $\lambda_3$ , are real, distinct, $\lambda_1\lambda_j > 0$ , $j = 2, 3$ and the eigenspaces are transverse to $M$ at 0;                            |
| <b>(a<sub>12</sub>)</b> Saddle M-Singularity:      | $X(0) = 0$ , the eigenvalues of $DX(0)$ , $\lambda_1, \lambda_2$ and $\lambda_3$ , are real, distinct, $\lambda_1\lambda_j < 0$ , $j = 2$ or $3$ and the eigenspaces are transverse to $M$ at 0;                          |
| <b>(a<sub>13</sub>)</b> Focal M-Singularity:       | 0 is a hyperbolic critical point of $X$ , the eigenvalues of $DX(0)$ are $\lambda_{12} = a \pm ib$ , $\lambda_3 = c$ , with $a, b, c$ distinct from zero and $c \neq a$ , and the eigenspaces are transverse to $M$ at 0. |
| <b>(b<sub>11</sub>)</b> Lips M-Singularity:        | presented in Definition 8, item 1, when $\text{Hess}(Fh _S(0)) > 0$ ;                                                                                                                                                     |
| <b>(b<sub>12</sub>)</b> Bec to Bec M-Singularity:  | presented in Definition 8, item 1, when $\text{Hess}(Fh _S(0)) < 0$ ;                                                                                                                                                     |
| <b>(b<sub>13</sub>)</b> Dove's Tail M-Singularity: | presented in <b>Definition 8, item 2</b> .                                                                                                                                                                                |

The next result is proved in (Medrado and Teixeira 1998). It deals with the normal forms of a codimension-one singularity.

**Theorem 13** i) (Generic Bifurcation and normal forms) Let  $X \in \chi(3)$ . The vector field  $X$  is  $M$ -structurally stable relative to  $\chi_1(3)$  if and only if  $X \in \Gamma_1(3)$ . ii) (Versal unfolding) In the space of one-parameter families of vector fields  $X_\alpha$  in  $\chi(3)$ ,  $\alpha \in (-\varepsilon, \varepsilon)$  an everywhere dense set is formed by generic families such that their normal forms are:

- $X_\alpha \in \Gamma_0(3)$ 
  - 0.1:  $X_\alpha(x, y, z) = (0, 0, 1)$
  - 0.2:  $X_\alpha(x, y, z) = (z, 0, \pm x)$
  - 0.3:  $X_\alpha(x, y, z) = (z, 0, x^2 + y)$
- $X_0 \in \Gamma_1(3)$ 
  - 1.1:  $X_\alpha(x, y, z) = \left(z, 0, \frac{-3x^2 + y^2 + \alpha}{2}\right)$
  - 1.2:  $X_\alpha(x, y, z) = \left(z, 0, \frac{-3x^2 - y^2 + \alpha}{2}\right)$
  - 1.3:  $X_\alpha(x, y, z) = \left(z, 0, \frac{4\delta x^3 + y + \alpha x}{2}\right)$ , with  $\delta = \pm 1$
  - 1.4:  $X_\alpha(x, y, z) = \left(axz, byz, \frac{ax + by + cz^2 + \alpha}{2}\right)$ , with  $(a, b, c) = \delta(3, 2, 1)$ ,  $\delta = \pm 1$
  - 1.5:  $X_\alpha(x, y, z) = \left(axz, byz, \frac{ax + by + cz^2 + \alpha}{2}\right)$ , with  $(a, b, c) = \delta(1, 3, 2)$ ,  $\delta = \pm 1$
  - 1.6:  $X_\alpha(x, y, z) = \left(axz, byz, \frac{ax + by + cz^2 + \alpha}{2}\right)$ , with  $(a, b, c) = \delta(1, 2, 3)$ ,  $\delta = \pm 1$
  - 1.7:  $X_\alpha(x, y, z) = \left(xz, 2yz, \frac{x + 2y - cz^2 + \alpha}{2}\right)$
  - 1.8:  $X_\alpha(x, y, z) = \left((-x + y)z, (-x - y)z, \frac{-3x - y + z^2 + \alpha}{2}\right)$

### Generic Bifurcation

Let  $Z = (X, Y) \in \Omega^r(n+1)$ . Call by  $\Sigma_0(n+1)$  (resp.  $\Sigma_1(n+1)$ ) the set of all elements that are structurally stable in  $\Omega^r(n+1)$  (resp.  $\Omega_1^r(n+1) = \Omega^r(n+1) \setminus \Sigma_0(n+1)$ ) in  $\Omega^r(n+1)$ . It is clear that a pre-classification of the generic singularities is immediately reached by:

If  $Z = (X, Y) \in \Sigma_0(n+1)$  (resp.  $Z = (X, Y) \in \Sigma_1(n+1)$ ) then  $X$  and  $Y$  are in  $\Gamma_0(n+1)$  (resp.  $X \in \Gamma_0(n+1)$  and  $Y \in \Gamma_1(n+1)$  or vice versa). Of course, the case when both  $X$  and  $Y$  are in  $\Gamma_1(n+1)$  is a codimension-two phenomenon.



## Two-Dimensional Case

The following result characterizes the structural stability in  $\Omega^r(2)$ .

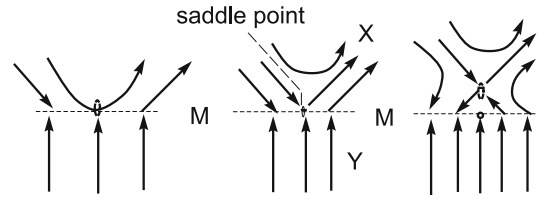
**Theorem A** (see Kozlova 1984; Sotomayor and Teixeira 1996):  $\Sigma_0(2)$  is an open and dense set of  $\Omega^r(2)$ . The vector field  $Z = (X, Y)$  is in  $\Sigma_0(2)$  if and only one of the following conditions is satisfied:

- i) Both elements  $X$  and  $Y$  are regular. When  $0 \in M$  is a simple singularity of  $Z$  then we assume that it is a hyperbolic critical point of  $Z^M$ .
- ii)  $X$  is a fold singularity and  $Y$  is regular (and vice-versa).

The following result still deserves a systematic proof. Following the same strategy stipulated in the generic classification of an  $M$ -singularity, Theorem 11 could be very useful. It is worthwhile to mention (Kuznetsov et al. 2003) where the problem of generic bifurcation in 2D was also addressed.

**Theorem B** (Generic Bifurcation) (see Machado and Sotomayor 2002; Sotomayor and Teixeira 1988)  $\Sigma_1(2)$  is an open and dense set of  $\Omega_1^r(2)$ . The vector field  $Z = (X, Y)$  is in  $\Sigma_1(2)$  provided that one of the following conditions is satisfied:

- i) Both elements  $X$  and  $Y$  are  $M$ -regular. When  $0 \in M$  is a simple singularity of  $Z$  then we assume that it is a codimension-one critical point (saddle-node or a Bogdanov–Takens singularity) of  $Z^M$ .
- ii)  $0$  is a codimension-one  $M$ -singularity of  $X$  and  $Y$  is  $M$ -regular. This case includes when  $0$  is either a cusp  $M$ -singularity or a critical point. Figure 4 illustrates the case when  $0$  is a saddle critical point in the boundary.
- iii) Both  $X$  and  $Y$  are fold  $M$ -singularities at  $0$ . In this case we have to impose that  $0$  is a hyperbolic critical point of the  $C^r$ -extension of  $Z^M$  provided that it is in the boundary of  $M_2 \cup M_3$  (see example below). Moreover when  $0$  is a distinguished *fold-fold* singularity of  $Z$  then  $0$  is a hyperbolic fixed point of the first return mapping  $\beta_Z$ .



**Perturbation Theory for Non-smooth Systems, Fig. 4**  $M$ -critical point for  $X$ ,  $M$ -regular for  $Y$  and its unfolding

Consider in a small neighborhood of  $0$  in  $R^2$ , the system  $Z = (X, Y)$  with  $X(x, y) = (1 - x^3 + y^2, x)$ ,  $Y(x, y) = (1 + x + y, -x + x^2)$  and  $h(x, y) = y$ . The point  $0$  is a *fold-fold*-singularity of  $Z$  with  $M_2 = \{x < 0\}$  and  $Z^s(x, 0) = (2x - x^2)^{-1}(2x - x^4 + x^5)$ . Observe that  $0$  is a hyperbolic critical point of the extended system  $G(x, y) = 2x - x^4 + x^5$ .

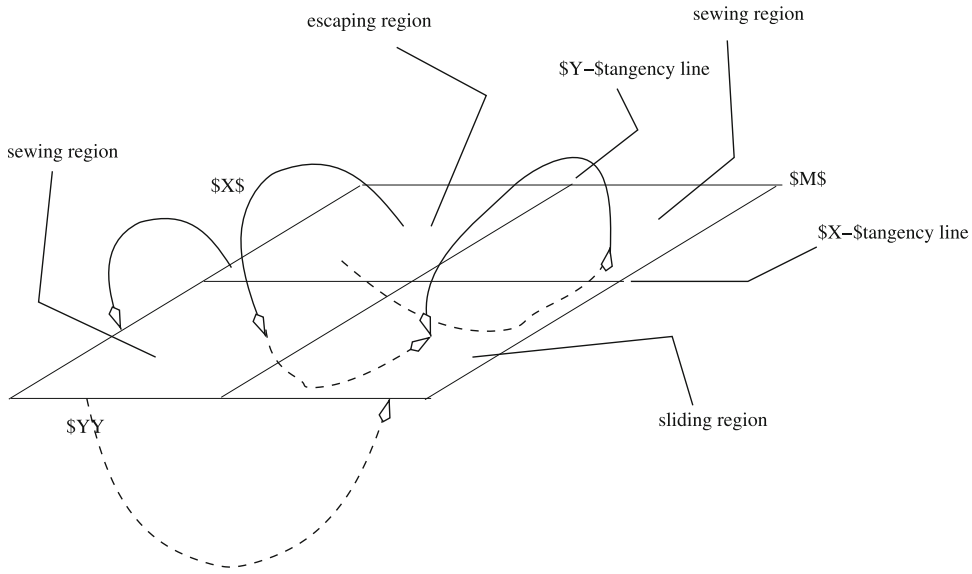
The classification of the codimension-two singularities in  $\Omega^r(2)$  is still an open problem. In this direction (Teixeira 1997) contains information about the classification of codimension-two  $M$ -singularities.

## Three-Dimensional Case

Let  $Z = (X, Y) \in \Omega^r(3)$ .

The most interesting case to be analyzed is when both vector fields,  $X$  and  $Y$  are fold singularities at  $0$  and the tangency sets  $\tau_X$  and  $\tau_Y$  in  $M$  are in general position at  $0$ . In fact they determine (in  $M$ ) four quadrants, two of them are  $M_1$ -regions, one is an  $M_3$ -region and the other is an  $M_2$ -region (see Fig. 5). We emphasize that the sliding vector field  $Z^M$  can be  $C^r$ -extended to a full neighborhood of  $0$  in  $M$ . Moreover,  $Z^M(0) = 0$ . Inside this class the *distinguished fold-fold* singularity (as defined in subsection “A Construction”) must be taken into account. Denote by the set of all *distinguished fold-fold singularities*  $Z \in \Omega^r(3)$ . Moreover, the eigenvalues of  $D\beta_Z(0)$  are  $\lambda = a \pm \sqrt{a^2 - 1}$ . If  $\lambda \in R$  we say that  $Z$  belongs to  $A_s$ . Otherwise  $Z$  is in  $A_e$ . Recall that  $\beta_Z$  is the first return mapping associated to  $Z$  and  $M$  at  $0$  as defined in subsection “A Construction”.

It is evident that the elements in the open set  $A_e$  are structurally unstable in  $\Omega^r(3)$ . It is worthwhile to mention that in  $A_e$  we detect elements which are



**Perturbation Theory for Non-smooth Systems, Fig. 5** The distinguished fold-fold singularity

asymptotically stable at the origin (Teixeira 1990). Concerning  $A_s$  few things are known.

We have the following result:

**Theorem C** The vector field  $Z = (X, Y)$  belongs to  $\Sigma_0(3)$  provided that one of the following conditions occurs:

- i) Both elements  $X$  and  $Y$  are regular. When  $0 \in M$  is a simple singularity of  $Z$  then we assume that it is a hyperbolic critical point of  $Z^M$ .
- ii)  $X$  is a fold singularity at  $0$  and  $Y$  is regular.
- iii)  $X$  is a cusp singularity at  $0$  and  $Y$  is regular.
- iv) Both systems  $X$  and  $Y$  are of fold type at  $0$ . Moreover: a) the tangency sets  $\tau_X$  and  $\tau_Y$  are in general position at  $0$  in  $M$ ; b) The eigenspaces associated with  $Z^M$  are transverse to  $\tau_X$  and  $\tau_Y$  at  $0 \in M$  and c)  $Z$  is not in  $A$ . Moreover the real parts of non conjugate eigenvalues are distinct.

We recall that bifurcation diagrams of sliding vector fields are presented in (Teixeira 1993; Teixeira 1999).

### Singular Perturbation Problem in 2D

Geometric singular perturbation theory is an important tool in the field of continuous dynamical systems. Needless to say that in this area very good

surveys are available (refer to Dumortier and Roussarie 1996; Fenichel 1979; Jones 1995). Here we highlight some results (see Buzzi et al. 2006) that bridge the space between discontinuous systems in  $\Omega^r(2)$  and singularly perturbed smooth systems.

**Definition 14** Let  $U \subset \mathbb{R}^2$  be an open subset and  $\varepsilon \geq 0$ . A singular perturbation problem in  $U$  (SP-Problem) is a differential system which can be written as

$$x' = \frac{dx}{d\tau} = f(x, y, \varepsilon), \quad y' = \frac{dy}{d\tau} = \varepsilon g(x, y, \varepsilon) \quad (6)$$

or equivalently, after the time re-scaling  $t = \varepsilon \tau$

$$\varepsilon \dot{x} = \varepsilon \frac{dx}{dt} = f(x, y, \varepsilon), \quad \dot{y} = \frac{dy}{dt} = g(x, y, \varepsilon), \quad (7)$$

with  $(x, y) \in U$  and  $f, g$  smooth in all variables.

Our first result is concerned with the transition between non-smooth systems and GSPT.

**Theorem D** Consider  $Z \in \Omega^r(2)$ ,  $Z_\varepsilon$  its  $\varphi$ -regularization, and  $p \in M$ . Suppose that  $\varphi$  is a

polynomial of degree  $k$  in a small interval  $I \subseteq (-1, 1)$  with  $0 \in I$ . Then the trajectories of  $Z_\varepsilon$  in  $V_\varepsilon = \{q \in R^2, 0 : h(q)/\varepsilon \in I\}$  are in correspondence with the solutions of an ordinary differential equation  $z' = \alpha(z, \varepsilon)$ , satisfying that  $\alpha$  is smooth in both variables and  $\alpha(z, 0) = 0$  for any  $z \in M$ . Moreover, if  $((X - Y)h^k)(p) \neq 0$  then we can take a  $C^{r-1}$ -local coordinate system  $\{(\partial/\partial x)(p), (\partial/\partial y)(p)\}$  such that this smooth ordinary differential equation is a SP-problem.

The understanding of the phase portrait of the vector field associated to a SP-problem is the main goal of the *geometric singular perturbation-theory* (GSP-theory). The techniques of GSP-theory can be used to obtain information on the dynamics of (6) for small values of  $\varepsilon > 0$ , mainly in searching minimal sets.

System (6) is called the *fast system*, and (7) the *slow system* of the SP-problem. Observe that for  $\varepsilon > 0$  the phase portraits of the fast and the slow systems coincide.

Theorem D says that we can transform a discontinuous vector field in a SP-problem. In general this transition cannot be done explicitly. Theorem E provides an explicit formula of the SP-problem for a suitable class of vector fields. Before the statement of such a result we need to present some preliminaries.

Consider  $C = \{\xi : R^2, 0 \rightarrow R\}$  with  $\xi \in C^r$  and  $L(\xi) = 0$  where  $L(\xi)$  denotes the linear part of  $\xi$  at  $(0, 0)$ .

Let  $\Omega_d \subset \Omega^r(2)$  be the set of vector fields  $Z = (X, Y)$  in  $\Omega^r(2)$  such that there exists  $\xi \in C$  that is a solution of

$$\nabla \xi(X - Y) = \Pi_i(X - Y), \quad (8)$$

where  $\nabla \xi$  is the gradient of the function and  $\Pi_i$  denote the canonical projections, for  $i = 1$  or  $i = 2$ .

**Theorem E** Consider  $Z \in \Omega_d$  and  $Z_\varepsilon$  its  $\varphi$ -regularization. Suppose that  $\varphi$  is a polynomial of degree  $k$  in a small interval  $I \subset R$  with  $0 \in I$ . Then the trajectories of  $Z_\varepsilon$  on  $V_\varepsilon = \{q \in R^2, 0 : h(q)/\varepsilon \in I\}$  are solutions of a SP-problem.

We remark that the singular problems discussed in the previous theorems, when  $\varepsilon \searrow 0$ ,

defines a dynamical system on the discontinuous set of the original problem. This fact can be very useful for problems in Control Theory.

Our third theorem says how the fast and the slow systems approximate the discontinuous vector field. Moreover, we can deduce from the proof that whereas the fast system approximates the discontinuous vector field, the slow system approaches the corresponding sliding vector field.

Consider  $Z \in \Omega^r(2)$  and  $\rho : R^2, 0 \rightarrow R$  with  $\rho(x, y)$  being the distance between  $(x, y)$  and  $M$ . We denote by  $\hat{Z}$  the vector field given by  $\hat{Z}(x, y) = \rho(x, y)Z(x, y)$ .

In what follows we identify  $\hat{Z}_\varepsilon$  and the vector field on  $\{R^2, 0\} \setminus M \times R\} \subset R^3$  given by  $\hat{Z}(x, y, \varepsilon) = (\hat{Z}_\varepsilon(x, y), 0)$ .

**Theorem F** Consider  $p = 0 \in M$ . Then there exists an open set  $U \subset R^2$ ,  $p \in U$ , a three-dimensional manifold  $M$ , a smooth function  $\Phi : M \rightarrow R^3$  and a SP-problem  $W$  on  $M$  such that  $\Phi$  sends orbits of  $W|_{\Phi^{-1}(U \times (0, +\infty))}$  in orbits of  $\hat{Z}|_{(U \times (0, +\infty))}$ .

### Examples

1. Take  $X(x, y) = (1, x)$ ,  $Y(x, y) = (-1, -3x)$ , and  $h(x, y) = y$ . The discontinuity set is  $\{(x, 0) \mid x \in R\}$ . We have  $Xh = x$ ,  $Yh = -3x$ , and then the unique non-regular point is  $(0, 0)$ . In this case we may apply Theorem E.
2. Let  $Z_\varepsilon(x, y) = (y/\varepsilon, 2xy/\varepsilon - x)$ . The associated partial differential equation (refer to Theorem E) with  $i = 2$  given above becomes  $2(\partial \xi / \partial x) + 4x(\partial \xi / \partial y) = 4x$ . We take the coordinate change  $\bar{x} = x, \bar{y} = y - x^2$ . The trajectories of  $X_\varepsilon$  in these coordinates are the solutions of the singular system

$$\varepsilon \dot{\bar{x}} = \bar{y} + \bar{x}^2, \quad \dot{\bar{y}} = -\bar{x}.$$

In what follows we try, by means of an example, to present a rough idea on the transition from non-smooth systems to GSPT. Consider  $X(x, y) = (3y^2 - y - 2, 1)$ ,  $Y(x, y) = (-3y^2 - y + 2, -1)$  and  $h(x, y) = x$ . The regularized vector field is

$$Z_\varepsilon(x, y) = \left( \frac{1}{2} + \frac{1}{2} \varphi\left(\frac{x}{\varepsilon}\right) \right) (3y^2 - y - 2, 1) \\ + \left( \frac{1}{2} - \frac{1}{2} \varphi\left(\frac{x}{\varepsilon}\right) \right) (-3y^2 - y + 2, -1).$$

After performing the polar blow up coordinates  $\alpha : [0, +\infty) \times [0, \pi] \times R \rightarrow R^3$  given by  $x = r \cos \theta$  and  $\varepsilon = r \sin \theta$  the last system is expressed by:

$$r\dot{\theta} = -\sin \theta (-y + \varphi(\cot \theta)(3y^2 - 2)), \\ \dot{y} = \varphi(\cot \theta).$$

So the slow manifold is given implicitly by  $\varphi(\cot \theta) = y/(3y^2 - 2)$  which defines two functions  $y_1(\theta) = \left(1 + \sqrt{1 + 24\varphi^2(\cot \theta)}\right)/(6\varphi(\cot \theta))$  and  $y_2(\theta) = \left(1 - \sqrt{1 + 24\varphi^2(\cot \theta)}\right)/(6\varphi(\cot \theta))$ . The function  $y_1(\theta)$  is increasing,  $y_1(0) = 1$ ,  $\lim_{\theta \rightarrow \pi/2^-} y_1(\theta) = +\infty$ ,  $\lim_{\theta \rightarrow \pi/2^+} y_1(\theta) = -\infty$  and  $y_1(\pi) = -1$ . The function  $y_2(\theta)$  is increasing,  $y_2(0) = -2/3$ ,  $\lim_{\theta \rightarrow \pi/2^-} y_2(\theta) = 0$  and  $y_2(\pi) = 2/3$ . We can extend  $y_2$  to  $(0, \pi)$  as a differential function with  $y_2(\pi/2) = 0$ .

The fast vector field is  $(\theta', 0)$  with  $\theta' > 0$  if  $(\theta, y)$  belongs to

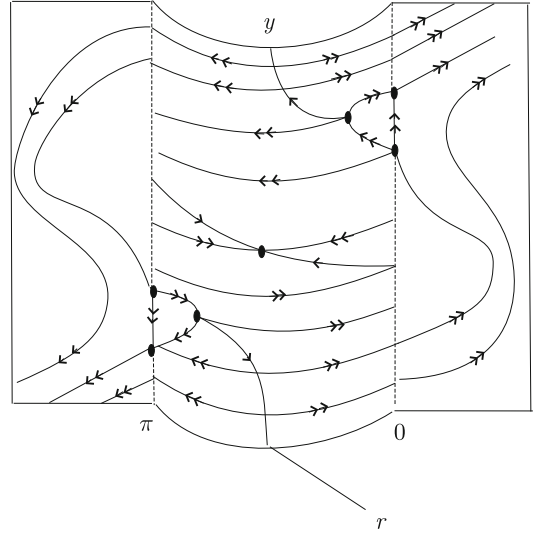
$$\left[ \left(0, \frac{\pi}{2}\right) \times (y_2(\theta), y_1(\theta)) \cup \left(\frac{\pi}{2}, \pi\right) \times (y_1(\theta), y_2(\theta)) \right] \\ \times (y_2(\theta), +\infty) \cup \left(\frac{\pi}{2}, \pi\right) \times (-\infty, y_1(\theta))$$

and with  $\theta' < 0$  if  $(\theta, y)$  belongs to

$$\left[ \left(0, \frac{\pi}{2}\right) \times (y_1(\theta), +\infty) \cup \left(0, \frac{\pi}{2}\right) \times (-\infty, y_2(\theta)) \cup \left(\frac{\pi}{2}, \pi\right) \times (y_1(\theta), y_2(\theta)) \right].$$

The reduced flow has one singular point at  $(0, 0)$  and it takes the positive direction of the  $y$ -axis if  $y \in (-\frac{2}{3}, 0) \cup (1, \infty)$  and the negative direction of the  $y$ -axis if  $y \in (-\infty, -1) \cup (0, \frac{2}{3})$ .

One can see that the singularities  $(\theta, y, r) = (0, 1, 0)$  and  $(\theta, y, r) = (0, -1, 0)$  are not normally



**Perturbation Theory for Non-smooth Systems, Fig. 6** Example of fast and slow dynamics of the SP-Problem

hyperbolic points. In this way, as usual, we perform additional blow ups. In Fig. 6 we illustrate the fast and the slow dynamics of the SP-problem. We present a phase portrait on the blowing up locus where a double arrow over a trajectory means that the trajectory belongs to the fast dynamical system, and a simple arrow means that the trajectory belongs to the slow dynamical system.

## Future Directions

Our concluding section is devoted to an outlook. Firstly we present some open problems linked with the setting that point out future directions of research. The main task for the future seems to bring the theory of non-smooth dynamical systems to a similar maturity as that of smooth systems. Finally we briefly discuss the main results in this text.

## Some Problems

In connection to this present work, some theoretical problems remain open:

The description of the bifurcation diagram of the codimension-two singularities in  $\Omega(2)$ . In this

last class we find some models (see Manosas and Torres 2005) where the following questions can also be addressed. a) When is a typical singularity topologically equivalent to a regular center? b) How about the isochronicity of such a center? c) When does a polynomial perturbation of such a system in  $\Omega(2)$  produce limit cycles? The articles (Broucke et al. 2001; Coll et al. 2001; Garcia and Teixeira 2004; Gasull and Torregrosa 2003; Teixeira 1979) can be useful auxiliary references.

1. Let  $\Omega(N)$  be the set of all non-smooth vector fields on a two-dimensional compact manifold  $N$  having a codimension-one compact submanifold  $M$  as its discontinuity set. The problem is to study the global generic bifurcation in  $\Omega(N)$ . The articles (Kozlova 1984; Kuznetsov et al. 2003; Percell 1973; Teixeira 1977) can be useful auxiliary references.
2. Study of the bifurcation set in  $\Omega'(3)$ . The articles (Medrado and Teixeira 1998; Percell 1973; Sotomayor and Teixeira 1988; Teixeira 1993) can be useful auxiliary references.
3. Study of the dynamics of the distinguished *fold-fold* singularity in  $\Omega'(n+1)$ . The article (Teixeira 1990) can be a useful auxiliary reference.
4. In many applications examples of non-smooth systems where the discontinuities are located on algebraic varieties are available. For instance, consider the system  $\ddot{x} + x \text{sign}(x) + \text{sign}(\dot{x}) = 0$ . Motivated by such models we present the following problem. Let  $0$  be a non-degenerate critical point of a smooth mapping  $h : R^{n+1}, 0 \rightarrow R, 0$ . Let  $\Phi(n+1)$  be the space of all vector fields  $Z$  on  $R^{n+1}, 0$  defined in the same way as  $\Omega(n+1)$ . We propose the following. i) Classify the typical singularities in that space. ii) Analyze the elements of  $\Phi(2)$  by means of “regularization processes” and the methods of GSPT, similarly to section “Vector Fields near the Boundary”. The articles (Alexander and Seidman 1998; Alexander and Seidman 1999) can be very useful auxiliary references.
5. In (Jacquemard and Teixeira 2003a; Jacquemard and Teixeira 2005) classes of 4D-relay systems are considered. Conditions

for the existence of one-parameter families of periodic orbits terminating at typical singularities are provided. We propose to find conditions for the existence of such families for  $n$ -dimensional relay systems.

## Conclusion

In this paper we have presented a compact survey of the geometric/qualitative theoretical features of non-smooth dynamical systems. We feel that our survey illustrates that this field is still in its early stages but enjoying growing interest. Given the importance and the relevance of such a theme, we have pointed above some open questions and we remark that there is still a wide range of bifurcation problems to be tackled. A brief summary of the main results in the text is given below.

1. We firstly deal with two-dimensional non-smooth vector fields  $Z = (X, Y)$  defined around the origin in  $R^2$ , where the discontinuity set is concentrated on the line  $\{y = 0\}$ . The first task is to characterize those systems which are structurally stable. This characterization is a starting point with which to establish a bifurcation theory as indicated by the Thom–Smale program.
2. In higher dimension the problem becomes much more complicated. We have presented here sufficient conditions for the three-dimensional local structural stability. Any further investigation on bifurcation in this context must pass through a deep analysis of the so called *fold-fold* singularity.
3. We have established a bridge between discontinuous and singularly perturbed smooth systems. Many similarities between such systems were observed and a comparative study of the two categories is called for.

## Bibliography

### Primary Literature

- Alexander JC, Seidman TI (1998) Sliding modes in intersecting switching surfaces I: blending. *Houst J Math* 24(3):545–569

- Alexander JC, Seidman TI (1999) Sliding modes in intersecting switching surfaces II: hysteresis. *Houst J Math* 25:185–211
- Andronov A, Pontryagin S (1937) Structurally stable systems. *Dokl Akad Nauk SSSR* 14:247–250
- Andronov AA, Vitt AA, Khaikin SE (1966) *Theory of oscillators*. Dover, New York
- Anosov DV (1959) Stability of the equilibrium positions in relay systems. *Autom Remote Control* XX(2):135–149
- Arnold VI (1983) *Methods in the theory of ordinary differential equations*. Springer, New York
- Bazykin AD (1998) *Nonlinear dynamics of interacting populations*. World Scientific Publishing Co Inc., River-Edge
- Bonnard B, Chyba M (2000) Singular trajectories and their role in control theory. *Mathématiques and applications*, vol 40. Springer, Berlin
- Broucke ME, Pugh CC, Simic SN (2001) Structural stability of piecewise smooth systems. *Comput Appl Math* 20(1–2):51–89
- Buzzi C, Silva PR, Teixeira MA (2006) A singular approach to discontinuous vector fields on the plane. *J Differ Equ* 23:633–655
- Chillingworth DR (2002) J.(4-SHMP) Discontinuity geometry for an impact oscillator. *Dyn Syst* 17(4):389–420
- Chua LO (1992) The genesis of Chua's circuit. *AEU* 46:250
- Coll B, Gasull A, Prohens R (2001) Degenerate Hopf bifurcations in discontinuous planar systems. *J Math Anal Appl* 253(2):671–690
- di Bernardo M, Budd C, Champneys AR, Kowalczyk P, Nordmark AB, Olivar G, Piiroinen PT (2005) *Bifurcations in non-smooth dynamical systems*. Bristol Centre for Applied Nonlinear Mathematics, N. 2005-4, Bristol
- Dumortier F (1977) Singularities of vector fields. IMPA, Rio de Janeiro
- Dumortier F, Roussarie R (1996) Canard cycles and center manifolds. *Memoirs Am Mathematical Soc* 121:x+100
- Ekeland I (1977) Discontinuités de champs hamiltoniens et existence de solutions optimales en calcul des variations. *Inst Hautes Études Sci Publ Math* 47:5–32
- Fenichel N (1979) Geometric singular perturbation theory for ordinary differential equations. *J Differ Equ* 31:53–98
- Filippov AF (1988) *Differential equations with discontinuous righthand sides*. Kluwer Academic, Dordrecht
- Flügge-Lotz I (1953) *Discontinuous automatic control*. Princeton University, Princeton, pp vii+168
- García R, Teixeira MA (2004) Vector fields in manifolds with boundary and reversibility—an expository account. *Qual Theory Dyn Syst* 4:311–327
- Gasull A, Torregrosa J (2003) Center-focus problem for discontinuous planar differential equations. *Int J Bifurc Chaos Appl Sci Eng* 13(7):1755–1766
- Henry P (1972) Differential equations with discontinuous right-hand side for planning procedures. *J Econ Theory* 4:545–551
- Hogan S (1989) On the dynamics of rigid-block motion under harmonic forcing. *Proc R Soc Lond A* 425:441–476
- Ito T (1979) A Filippov solution of a system of diff. eq. with discontinuous right-hand sides. *Econ Lett* 4:349–354
- Jackson JD (1999) *Classical electrodynamics*, 3rd edn. Wiley, New York, pp xxii+808
- Jacquemard A, Teixeira MA (2003a) Computer analysis of periodic orbits of discontinuous vector fields. *J Symbol Comput* 35:617–636
- Jacquemard A, Teixeira MA (2003b) On singularities of discontinuous vector fields. *Bull Sci Math* 127:611–633
- Jacquemard A, Teixeira MA (2005) Invariant varieties of discontinuous vector fields. *Nonlinearity* 18:21–43
- Jones C 1995 *Geometric singular perturbation theory*. C.I.M.E. Lectures, Montecatini Terme, June 1994, Lecture notes in mathematics 1609. Springer, Heidelberg
- Kozlova VS (1984) Roughness of a discontinuous system. *Vestnik Moskovskogo Universiteta, Matematika* 5:16–20
- Kunze M, Kopper T (1997) Qualitative bifurcation analysis of a non-smooth friction oscillator model. *Z Angew Math Phys* 48:87–101
- Kuznetsov YA et al (2003) One-parameter bifurcations in planar Filippov systems. *Int J Bifurc Chaos* 13:2157–2188
- Llibre J, Teixeira MA (1997) Regularization of discontinuous vector fields in dimension 3. *Discrete Continuous Dyn Syst* 3(2):235–241
- Luo CJ (2006) *Singularity and dynamics on discontinuous vector fields*. Monograph series on nonlinear science and complexity. Elsevier, New York, pp i+310
- Machado AL, Sotomayor J (2002) Structurally stable discontinuous vector fields in the plane. *Qual Theory Dyn Syst* 3:227–250
- Manosas F, Torres PJ (2005) Isochronicity of a class of piecewise continuous oscillators. *Proc AMS* 133(10):3027–3035
- Medrado J, Teixeira MA (1998) Symmetric singularities of reversible vector fields in dimension three. *Phys D* 112:122–131
- Medrado J, Teixeira MA (2001) Codimension-two singularities of reversible vector fields in 3D. *Qual Theory Dyn Syst J* 2(2):399–428
- Percell PB (1973) Structural stability on manifolds with boundary. *Topology* 12:123–144
- Seidman T (2006) Aspects of modeling with discontinuities. In: N'Guerekata G (ed) *Advances in applied and computational mathematics*, Proceedings of the Dover conference. Cambridge. <http://www.umbc.edu/%7Eseidman>
- Sotomayor J (1974) Structural stability in manifolds with boundary. In: *Global analysis and its applications*, vol III. IEAA, Vienna, pp 167–176
- Sotomayor J, Teixeira MA (1988) *Vector fields near the boundary of a 3-manifold*. Lecture notes in mathematics, vol 331. Springer, Berlin/Heidelberg, pp 169–195



- Sotomayor J, Teixeira MA (1996) Regularization of discontinuous vector fields. International Conference on Differential Equations, Lisboa. World Scientific, Singapore, pp 207–223
- Sussmann H (1979) Subanalytic sets and feedback control. J Differ Equ 31:31–52
- Teixeira MA (1977) Generic bifurcations in manifolds with boundary. J Differ Equ 25:65–89
- Teixeira MA (1979) Generic bifurcation of certain singularities. Boll Unione Mat Ital 16-B:238–254
- Teixeira MA (1990) Stability conditions for discontinuous vector fields. J Differ Equ 88:15–24
- Teixeira MA (1991) Generic singularities of discontinuous vector fields. An Ac Bras Cienc 53(2):257–260
- Teixeira MA (1993) Generic bifurcation of sliding vector fields. J Math Anal Appl 176:436–457
- Teixeira MA (1997) Singularities of reversible vector fields. Phys D 100:101–118
- Teixeira MA (1999) Codimension-two singularities of sliding vector fields. Bull Belg Math Soc 6(3):369–381
- Vishik SM (1972) Vector fields near the boundary of a manifold. Vestnik Moskovskogo Universiteta, Matematika 27(1):13–19
- Zelikin MI, Borisov VF (1994) Theory of chattering control with applications to astronautics, robotics, economics, and engineering. Birkhäuser, Boston
- Huertas JL, Chen WK, Madan RN (eds) (1997) Visions of nonlinear science in the 21st century, Part I. Festschrift dedicated to Leon O. Chua on the occasion of his 60th birthday. Papers from the workshop held in Sevilla, June 26, 1996. Int J Bifurc Chaos Appl Sci Eng 7 (1997), no. 9. World Scientific Publishing Co. Pte. Ltd., Singapore, pp i–iv, pp 1907–2173
- Komuro M (1990) Periodic bifurcation of continuous piecewise vector fields. In: Shiraiwa K (ed) Advanced series in dynamical systems, vol 9. World Scientific, Singapore
- Kunze M (2000) Non-smooth dynamical systems. Lecture notes in mathematics, vol 1744. Springer, Berlin, pp x+228
- Kunze M, Küpper T (2001) Non-smooth dynamical systems: an overview. In: Fiedler B (ed) Ergodic theory, analysis, and efficient simulation of dynamical systems. Springer, Berlin, pp 431–452
- Libre J, Silva PR, Teixeira MA (2007) Regularization of discontinuous vector fields on  $R^3$  via singular perturbation. J Dyn Differ Equ 19(2):309–331
- Martinet J (1974) Singularités des fonctions et applications différentiables. Pontificia Universidade Católica do Rio de Janeiro, Rio de Janeiro, pp xiii+181 (French)
- Minorsky N (1962) Nonlinear oscillations. D Van Nostrand Co Inc, Princeton/Toronto/London/New York, pp xviii+714
- Minorsky N (1967) Théorie des oscillations. Mémorial des Sciences Mathématiques Fasc, vol 163. Gauthier-Villars, Paris, pp i+114 (French)
- Minorsky N (1969) Theory of nonlinear control systems. McGraw-Hill, New York/London/Sydney, pp xx+331
- Ostrowski JP, Burdick JW (1997) Controllability for mechanical systems with symmetries and constraints. Recent developments in robotics. Appl Math Comput Sci 7(2):305–331
- Peixoto MC, Peixoto MM (1959) Structural stability in the plane with enlarged conditions. Anais da Acad Bras Ciências 31:135–160
- Rega G, Lenci S (2003) Nonsmooth dynamics, bifurcation and control in an impact system. Syst Anal Model Simul Arch 43(3), Gordon and Breach Science Publishers, Inc. Newark, 343–360
- Seidman T (1995) Some limit problems for relays. In: Lakshmikantham V (ed) Proceedings of the first world congress of nonlinear analysis, vol I. Walter de Gruyter, Berlin, pp 787–796
- Sotomayor J (1974) Generic one-parameter families of vector fields on 2-manifolds. Publ Math IHES 43:5–46
- Szmolyan P (1991) Transversal heteroclinic and homoclinic orbits in singular perturbation problems. J Differ Equ 92:252–281
- Teixeira MA (1985) Topological invariant for discontinuous vector fields. Nonlinear Anal TMA 9(10): 1073–1080
- Utkin V (1978) Sliding modes and their application in variable structure systems. Mir, Moscow
- Utkin V (1992) Sliding modes in control and optimization. Springer, Berlin

## Books and Reviews

- Agrachev AA, Sachkov YL (2004) Control theory from the geometric viewpoint. Encyclopaedia of mathematical sciences, 87. Control theory and optimization, vol II. Springer, Berlin, pp xiv+412
- Barbashin EA (1970) Introduction to the theory of stability. Wolters-Noordhoff Publishing, Groningen. Translated from the Russian by Transcripts Service, London. Edited by T. Lukes, p 223
- Brogliato B (ed) (2000) Impacts in mechanical systems. Lecture notes in physics, vol 551. Springer, Berlin, p 160
- Carmona V, Freire E, Ponce E, Torres F (2005) Bifurcation of invariant cones in piecewise linear homogeneous systems. Int J Bifurc Chaos 15(8):2469–2484
- Davydov AA (1994) Qualitative theory of control systems. American Mathematical Society, Providence, RI. (English summary) Translated from the Russian manuscript by V. M. Volosov. Translations of Mathematical Monographs, 141, pp viii+147
- Dercole F, Gragnani F, Kuznetsov YA, Rinaldi S (2003) Numerical sliding bifurcation analysis: an application to a relay control system. IEEE Trans Circuit Systems – I: Fund Theory Appl 50:1058–1063
- Glocker C (2001) Set-valued force laws: dynamics of non-smooth systems. Lecture notes in applied mechanics, vol 1. Springer, Berlin
- Hogan J, Champneys A, Krauskopf B, di Bernardo M, Wilson E, Osinga H, Homer M (eds) (2003) Nonlinear dynamics and chaos: where do we go from here? Institute of Physics Publishing (IOP), Bristol, pp xi+358

# Exact and Perturbation Methods in the Dynamics of Legged Locomotion

Oleg Makarenkov

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA

## Article Outline

Glossary

Definition of the Subject

Introduction

Poincaré Maps for Systems with Impacts

The Planar Rimless Wheel (Poincaré Map Is Explicitly Computable)

The Actuated Planar Biped (Poincaré Map Reduces to an Explicitly Computable Map)

Fixed Points of Perturbed Poincaré Maps

Compass-Gait Biped (Poincaré Map Is a Perturbation of an Explicitly Computable Map)

Future Directions

Bibliography

## Glossary

**Border-collision bifurcation** For impact systems, *border-collision bifurcation* is the bifurcation that occurs when an equilibrium of an impact system collides with the impact surface under varying parameters.

**Compass-gait biped** A double pendulum consists of two links (*legs*) with point masses on them and a third point mass at the joint (*hip*) connecting the two links. The free ends of the legs are called *foots*. A *compass-gait biped* is a planar (i.e., 2D) double pendulum with legs of equal lengths placed on a line (*ground*) with *foots* down and *hip* up. The dynamics (*gait*) of a compass-gait biped is defined by three

parameters being the slope of the ground, the normalized mass, and the length of the legs. Due to the collisions of legs with the ground, the dynamics of a compass-gait biped is governed by an impact system. Periodic solutions of the corresponding impact system define walking cycles (*walking gaits* of the biped).

**Grazing bifurcation** For impact systems, *grazing bifurcation* is the bifurcation that occurs when a periodic solution of an impact system collides with the impact surface under varying parameters.

**Hybrid zero dynamics** For an impact system, *hybrid zero dynamics* is the dynamics on a stable manifold of the impact system.

**Impact system** It consists of a continuous dynamical system, a codimension one manifold (called *impact surface* or *switching manifold*), and an impact law, so that the trajectory of the dynamical system jumps according to the impact law each time it reaches the impact surface.

**Perturbation theory** It is an analytic tool that allows making local and global conclusions about the dynamics of continuous or discrete perturbed dynamical system based on fully computable dynamics of the unperturbed system.

**Poincaré map** Poincaré map of a continuous dynamical system is a map whose fixed points (and their stability properties) are linked to periodic solutions (and their stability properties) of the dynamical system. A Poincaré map is a discrete dynamical system.

## Definition of the Subject

Compass-gait biped is the simplest robot that walks similar to human (*anthropomorphically*) under the action of just gravity (i.e., without control or *passively*). In order to take the most from the dynamics of the compass-gait biped in the design of more complex actuated

(i.e., controlled) robotic walkers, an analytic framework that would systematically clarify the stability properties of the compass-gait biped is required. As a matter of fact, the Poincaré map of the compass-gait biped cannot be computed explicitly and perturbation theory is the only currently known rigorous method that succeeds in proving (analytically) the existence and stability of walking cycles in the compass-gait biped. Therefore, perturbation theory is a natural starting point for the required analytic framework to be built upon. The present survey proposes a program for such a development. The literature review aims to set up a background and to make the proposed program compelling.

## Introduction

The greatest energy efficiency of legged locomotion can be achieved by taking the most from the self-oscillating dynamics of *passive walkers* (Collins et al. 2005). Though there are various passivity-based strategies to control locomotion of both 2D- and 3D-legged bipeds (see Beigzadeh et al. 2018; Ikeda and Toyama 2015; Spong et al. 2007; Spong 1999; Wang et al. 2009; Wisse et al. 2005), the use of passivity is still limited because the dynamics of even the simplest two-link passive biped by McGeer (1990) (*compass-gait biped*) is not fully understood. Indeed, the walking cycle of McGeer's model has an extremely narrow domain of attraction (Kerimoglu et al. 2021; Liu et al. 2008; Obayashi et al. 2016; Schwab and Wisse 2001; Sidorov and Zacksenhouse 2019), and there is no systematic theory as for what ingredients are responsible for the size of this domain. The model commonly accepted in robotics literature as an adequate restriction of the McGeer's model is a so-called rimless wheel model, whose dynamics is just one-dimensional and whose Poincaré map can be computed explicitly (Asano 2015; Byl and Tedrake 2009; Coleman 2010). Rimless-wheel-like models with explicitly computable Poincaré maps are used in control strategies for the target walking cycles because the domains of attraction of such walking cycles are easily computable and often

global (Grizzle et al. 2001; Westervelt et al. 2003, 2004). The Poincaré map of compass-gait biped is two-dimensional and cannot be computed in closed form; see (Garcia et al. 1998; Gritli and Belghith 2016; Makarenkov 2020). The goal of the survey is to expose the reader to the so-called *perturbation methods* that are capable to analyze the dynamics of Poincaré maps that are not given explicitly with particular focus on compass-gait biped and its extensions. To embed the perturbation approach into the landscape of robotics research more naturally, necessary topics of robotics literature are surveyed as well.

Perturbation methods are capable to analyze the dynamics that cannot be computed in closed form (Bogoliubov and Mitropolsky 1961; Guckenheimer and Holmes 1990; Malkin 1956; Sanders et al. 2007). For the readers familiar with the implicit function theorem from a course of analysis, it would be quite correct to view perturbation theory as a generalization of the implicit function theorem from the space of vectors to the space of dynamical systems. Specifically, perturbation theory assumes that the dynamical system under consideration is close to such a system whose dynamics (called *reduced dynamics*) is computable in closed form. When this assumption holds, perturbation theory makes conclusions about the dynamics of the full system based on the properties of the reduced dynamics and perturbation terms.

The model of a bipedal walker is a so-called *impact system* that consists of differential equations and an impact law. Differential equations govern the dynamics of the biped between collisions of legs with ground (*heel strikes*). In between successive heel strikes, one of the legs (*stance leg*) holds a contact with the ground and another leg does not touch the ground (*swing leg*). The biped model between collisions can be viewed as a double pendulum (Goswami et al. 1998; McGeer 1990). The legs swap their roles when swing leg collides with the ground, at which point the trajectory jumps and the impact law describes how exactly the trajectory jumps. The impact law is derived from the law of conservation of momentums. Using the language of dynamical systems, the impact law applies when

the trajectory reaches a codimension one manifold (called *impact surface* or *switching manifold*). When equations of the biped and properties of the ground do not change from impact to impact, the model of the biped generally admits a limit cycle (Freidovich et al. 2009; Garcia et al. 1998; Grizzle et al. 2001; Makarenkov 2020; Westervelt et al. 2003) also called *walking limit cycle*. Locating the walking limit cycle and improving its properties is one of the central problems in robotic locomotion.

Pioneering research on impact systems goes back to the textbook (Andronov et al. 1987), where the existence and stability of limit cycles was established for models of clock. The analysis in Andronov et al. (1987) is based on setting up a Poincaré map whose formula turned out to be computable explicitly thanks to availability of the integral of motion. General methods to study the existence of attracting limit cycles whose Poincaré maps are not computable explicitly have been developed over the recent 30 years based on bifurcation and perturbation theories. The two central mechanisms of how an attracting limit cycle with jumps can occur are border-collision and grazing bifurcations (di Bernardo et al. 2008a, b; Gardini et al. 2014; Glendinning 2015, 2016; Guardia et al. 2011; Kowalczyk et al. 2006; Kupper and Moritz 2001; Leine and Nijmeijer 2004; Makarenkov and Niwanthi Wadippuli Achchige 2018; Makarenkov and Lamb 2012; Nordmark 1991; Simpson and Meiss 2007, 2012; Simpson 2010; Sushko and Gardini 2010; Weiss et al. 2012; Zou et al. 2006). Border-collision bifurcation of a limit cycle is an analogue of the classical Hopf bifurcation which makes sense for systems with switching manifolds. In the border-collision bifurcation of a limit cycle, the occurrence of a limit cycle with a jump is caused by a collision of an equilibrium with the impact surface under a varying parameter. In the grazing bifurcation, varying parameter causes displacement of a limit cycle without jumps toward the impact surface and the limit cycle gains jumps upon colliding with the impact surface. The latter scenario requires an additional assumption to presume stability of the limit cycle, which holds in some applications

(Misra et al. 2010; Rom-Kedar and Turaev 1999; Turaev and Rom-Kedar 1998; Yagasaki 2004; Zhao and Dankowicz 2006) but does not hold generically (di Bernardo et al. 2008a; Nordmark 1991). For more advance scenarios of how a limit cycle in systems with thresholds can appear (bifurcate) under varying parameter, the reader is referred to (Makarenkov 2017a, 2019; Simpson 2018) (border-splitting bifurcation) and (Castillo 2020; Cristiano et al. 2019) (bifurcation from Teixeira singularity).

The reader may wish to note that impact systems are often referred to as *hybrid systems* in robotics literature (Kolathaya and Ames 2017; Westervelt et al. 2003). A general hybrid system is represented by a family of systems of differential equations and by a graph which explains how the members of the family can switch between each other (DeCarlo et al. 2000; Heemels and Brogliato 2003; Simic et al. 2005; Ye et al. 1998). In particular, one system is allowed to switch to multiple other members of the graph or to itself, and each such switch is accompanied by its own jump law (called *reset*). The jump occurs when the trajectory reaches a switching manifold (described in hybrid systems over so-called *guards*). And, as just said, the trajectory is allowed to follow multiple jump laws depending on the system to which the trajectory “wants” to jump. Solutions of hybrid systems are called *executions*. Impact system is a particular case of hybrid system where each trajectory has only one option to jump when it reaches a switching manifold (see Goebel et al. (2009, 2012) for further reading on convenient reformulations of hybrid systems). The term *impact system* will be used in this survey from now on.

The survey is organized as follows. Before embarking in sections “Fixed Points of Perturbed Poincaré Maps” and “Compass-Gait Biped (Poincaré Map Is a Perturbation of an Explicitly Computable Map)” on a discussion of the perturbation theory and its application to compass-gait biped, section “The Actuated Planar Biped (Poincaré Map Reduces to an Explicitly Computable Map)” presents a so-called zero dynamics approach which explains how a passive two-link walker can be obtained as the limit of the dynamics

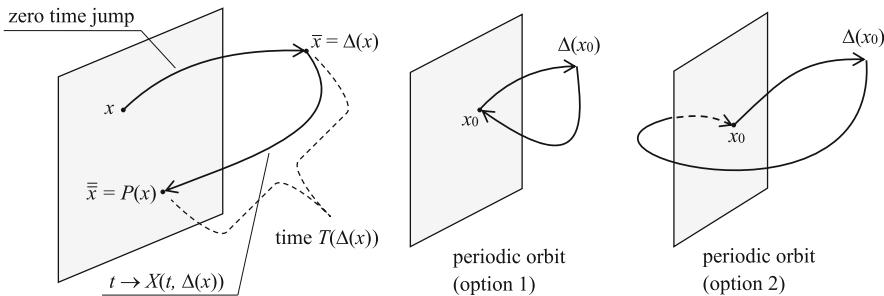
of a controlled multilink biped. Following the current literature, the passive walker obtained using zero dynamics approach in section “[The Actuated Planar Biped \(Poincaré Map Reduces to an Explicitly Computable Map\)](#)” is not compass-gait biped, but a simpler walker conceptually equivalent to the rolling rimless wheel. Section “[The Planar Rimless Wheel \(Poincaré Map Is Explicitly Computable\)](#)” shows that thanks to the presence of an explicit integral of motion (10), the analysis of the rimless wheel model reduces to the analysis of an explicit one-dimensional Poincaré map. The aim of section “[The Planar Rimless Wheel \(Poincaré Map Is Explicitly Computable\)](#)” is to familiarize the reader with the Poincaré map approach whose main tool (Theorem 1) is presented earlier in section “[Poincaré Maps for Systems with Impacts.](#)” When the Poincaré map is available explicitly, Theorem 1 (based on contraction-mapping principle) gives a simple answer about the existence and stability of fixed points (and, respectively, walking limit cycles). With this background in mind, the reader is then exposed to the equations of the compass-gait biped which can be viewed as the rimless wheel model where the angle between spokes can vary. The equations of passive biped no longer allow an explicit integral of motion, which brings us naturally to the use of perturbation theory for Poincaré maps presented in section “[Fixed Points of Perturbed Poincaré Maps](#)” (Theorem 2) and applied to simplified (according to Garcia et al. 1998) equations of compass-gait biped in section “[Compass-Gait Biped \(Poincaré Map Is a Perturbation of an Explicitly Computable Map\)](#).”

Building upon the notations of previous sections, a plan for further development of perturbation theory that aims to enlarge the domain of attraction of the walking cycle of passive bipeds is presented in section “[Future Directions.](#)” Possible directions to investigate nonperiodic motions of passive biped are also proposed. Section “[Future Directions](#)” focuses on locomotion in just the sagittal plane (with the slope playing the role of small parameter). However, some questions toward the locomotion in both sagittal and frontal directions that are natural consequences of the 2D discussion are touched upon as well (subsection “[The Effect of Accounting for the Third Dimension \(Higher-Dimensional Perturbation Analysis\)](#)”).

There is a good chance that the ideas of section “[Future Directions,](#)” if executed successfully, will provide engineers with an analytic framework toward the use of compass-gait-like bipeds as the models of zero dynamics (17). This will, in particular, reduce the cost of locomotion for those robots which run based on zero dynamics of the rimless wheel type. On the other hand, the survey intends to give a good deal of motivations to mathematicians working in the field of perturbation theory and interested in engineering applications.

## Poincaré Maps for Systems with Impacts

**Definition 1** A set  $S \subset \mathbb{R}^m$  is a  $C^1$ -manifold of codimension one, if



**Exact and Perturbation Methods in the Dynamics of Legged Locomotion, Fig. 1** Illustration of the notations of section “[Poincaré Maps for Systems with Impacts](#)”

$$S = \{x \in \overline{\Omega} : H(x) = 0\}, \quad (1)$$

for some  $H \in C^1(\overline{\Omega}, \mathbb{R})$  and  $H'(x) \neq 0$ ,  $x \in \Omega$ , where  $\Omega \in \mathbb{R}^m$  is an open set. A differential equation with impacts is a smooth differential equation

$$\dot{x} = f(x), \quad x \in \mathbb{R}^m, \quad (2)$$

coupled with an impact map  $\Delta: S \rightarrow \mathbb{R}^m$  which creates a discontinuity in solution  $x$  every time  $x$  reaches a codimension one  $C^1$  manifold  $S$  (called *switching manifold* or *impact surface*). Specifically, every time the relation  $x(t^-) \in S$  holds (i.e.  $\lim_{\tau \rightarrow t^-} x(\tau) \in S$ ),  $\Delta$  applies instantaneously and sets  $x(t)$  to the position  $\Delta(x(t^-))$ : The solution  $x$  then stays continuous until it reaches  $S$  again. This rule can be formulated as

$$\begin{aligned} \dot{x} &= f(x), \\ x(t^+) &= \Delta(x(t^-)), \quad \text{if } x(t^-) \in S. \end{aligned} \quad (3)$$

There is an uncertainty as for how to define a solution of Eq. (3) at the time of impact. One of the conventions is to view solutions of Eq. (3) right-continuous (Edmond and Thibault 2006; Zhai et al. 2001).

Let us denote by  $t \rightarrow X(t, x)$  the solution  $x(t)$  of Eq. (2) with the initial condition  $x(0) = x$ . Recall that according to Definition 1, a  $C^1$  smooth manifold  $S$  comes with a function  $H \in C^1(\mathbb{R}^m, \mathbb{R})$  such that Eq. (1) holds for some  $H \in C^1(\overline{\Omega}, \mathbb{R})$  and  $H'(x) \neq 0$ ,  $x \in \Omega$ , where  $\Omega \subset \mathbb{R}^m$  is an open set.

**Definition 2** If for  $x \in \mathbb{R}^m$  there exists  $T(x) > 0$  such that

$$\begin{aligned} H(X(T(x), x)) &= 0 \quad \text{and} \quad H(X(t, x)) \neq 0, \\ t &\in (0, T(x)), \end{aligned} \quad (4)$$

then the map

$$\mathcal{P}(x) = X(T(x), x) \quad (5)$$

is called the *Poincaré map* of system (2) induced by cross-section  $S$ . The map  $x \mapsto T(x)$  is called the *time map* of system (2).

If, for a given  $x \in \mathbb{R}^m$ , the value of  $T > 0$  demanded in Definition 2 does not exist, the *Poincaré map* of system (2) induced by cross-section  $S$  is said to be undefined at  $x$ .

**Definition 3** (Grizzle et al. 2001; Westervelt et al. 2003) The *Poincaré map*  $P$  of impact system (3) induced by cross-section  $S$  is defined as

$$P(x) = \mathcal{P}(\Delta(x)) \quad (6)$$

provided that  $\mathcal{P}$  is defined at  $\Delta(x)$ :

When used in connection to the impact system (3), the time map  $x \mapsto T(x)$  of Definition 2 is called *time to impact*. A fixed point  $x_0$  of Poincaré map  $P$  corresponds to the initial conditions of a cycle  $x_*(t)$  of impact system (3). The value of  $T(x_0)$  is the period of cycle  $x_*$ . A cycle is called a *limit cycle*, if the cycle is isolated, i.e., the fixed point  $x_0$  of  $P$  is isolated.

**Remark 1** A part of robotics literature (see the fundamental paper (Garcia et al. 1998) that is considered in section “Compass-Gait Biped (Poincaré Map Is a Perturbation of an Explicitly Computable Map)” in detail) defines the Poincaré map of Eq. (3) as

$$P(x) = \Delta(\mathcal{P}(x)). \quad (7)$$

The advantage of the latter definition is that it matches the right-continuity of the solution of Eq. (3) in the sense that if  $x(t)$  is a right-continuous solution of Eq. (3) then  $P(x(0)) = x(T(x(0)))$ . The advantage of Definition 3 is that it always keeps the impact surface  $S$  invariant. The map (7) keeps  $S$  invariant only if  $S$  is invariant for  $\Delta$  (which is quite often the case in robotics applications, see section “Compass-Gait Biped (Poincaré Map Is a Perturbation of an Explicitly Computable Map)”).

To define the notion of stability of cycle  $x_*$ , let us introduce the  $\varepsilon$ -neighborhood of cycle  $x_*$  as follows. Given  $\varepsilon > 0$ , consider  $U = B_\varepsilon(x_0) \cap S$ , where  $B_\varepsilon(x_0)$  is the open ball in  $\mathbb{R}^m$  centered at  $x_0$  and of radius  $\varepsilon$ . Assume that  $x \mapsto T(x)$  is defined on  $U$  and call



$$U_\varepsilon = \bigcup_{t \in [0, T(\xi)], \xi \in U} \{X(t, \xi)\}$$

the  $\varepsilon$ -neighborhood of cycle  $x_*$ . Adapting Andronov et al. (1987, Chap. 5, §6), the following definition of orbital stability of periodic cycle  $x_*$  can be formulated.

**Definition 4** A cycle  $x_*$  of period  $T$  of impact system (3) is an orbitally stable limit cycle, if there exists  $\varepsilon > 0$  such that, for any  $\xi \in U_\varepsilon$ , the solution  $x(t)$  of Eq. (3) with the initial condition  $x(0) = \xi$  approaches the set  $x_*([0, T))$  as  $t \rightarrow \infty$ . The corresponding set  $U_\varepsilon$  is the basin of attraction of  $x_*$ .

Note, Definition 4 is defined only if the time map of cycle  $x_*$  is defined at any point of  $S$  that is  $\varepsilon$ -close to  $x_*(0)$ .

**Theorem 1** If

- (i)  $\Delta : S \mapsto \Delta(S)$  is a contraction,
- (ii)  $\mathcal{P} : \Delta(S) \mapsto S$  is defined and is a contraction,

then the map  $P : S \rightarrow S$  given by Eq. (6) is a contraction (with some constant  $\alpha > 0$ ). In particular,  $P$  has a unique fixed point  $x_0$  in  $S$  such that

$$P(B_\varepsilon(x_0) \cap S) \subset B_{\alpha\varepsilon}(x_0) \cap S, \quad \varepsilon > 0,$$

and the corresponding limit cycle  $x_*$  of impact system (3) is orbitally stable with the basin of attraction

$$\bigcup_{t \in [0, T(\xi)], \xi \in S} \{X(t, \xi)\}.$$

The proof follows directly from Banach Contraction Principle and Definition 4 (the reader can consult Krasnoselskii (1968, §9).

The next section discusses the most typical for robotics literature application of Theorem 1.

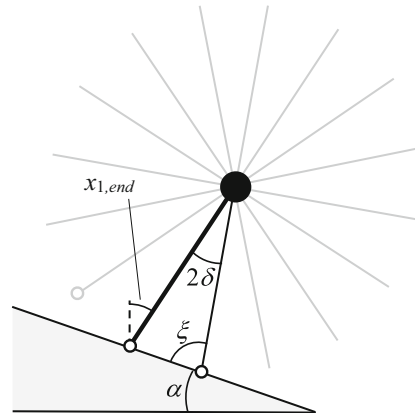
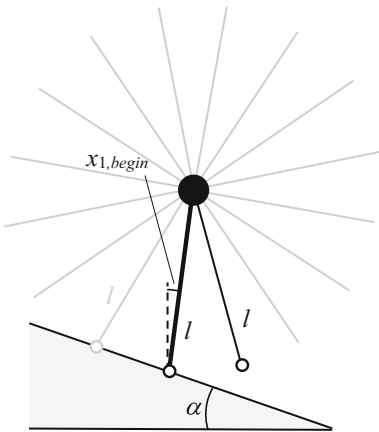
### The Planar Rimless Wheel (Poincaré Map Is Explicitly Computable)

Denoting by  $x_1$  the angle of the spoke in contact from vertical, the differential equation for the single spoke in contact reads as (Saglam et al. 2014)

$$\dot{x} = f(x) = \begin{pmatrix} x_2 \\ \zeta \sin(x_1) \end{pmatrix}, \quad (8)$$

where  $\zeta = g/l$ ,  $g$  is the acceleration due to gravity, and  $l$  is the length of the spokes.

From the triangles of Fig. 2, one computes (Clark and Bloch 2020)



**Exact and Perturbation Methods in the Dynamics of Legged Locomotion, Fig. 2** A phase of rimless wheel rolling between two successive collisions. The gray leg is stance leg until  $x_{1,begin}$  excluding  $x_{1,begin}$ . The

bold black leg is stance leg on  $[x_{1,begin}, x_{1,end}]$ . The regular black leg is stance leg beginning  $x_{1,end}$  (see Saglam et al. 2014, Fig. 1)

$$x_{1,begin} = \zeta + \alpha - \frac{\pi}{2}, \quad x_{1,end} = \zeta + \alpha - \frac{\pi}{2} + 2\delta, \\ \zeta = \arcsin(\sin(2\delta)/s), \quad s = \sqrt{2 - 2\cos(2\delta)}.$$

Therefore, the following quantities will be used

$$D = \left\{ (x_1, x_2) : \zeta + \alpha - \frac{\pi}{2} < x_1 < \zeta + \alpha - \frac{\pi}{2} \right. \\ \left. + 2\delta, x_2 \geq 0 \right\}, \\ S = \{(x_1, x_2) \in D : H(x) = 0\}, \\ H(x) = x_1 - \left( \zeta_1 + \alpha - \frac{\pi}{2} + 2\delta \right).$$

When a solution  $x(t)$  of Eq. (8) reaches  $S$ , a collision occurs which switches the stance

leg and amends  $x(t)$  according to the impact law

$$x(t) \rightarrow \Delta(x(t)) = \begin{pmatrix} \zeta + \alpha - \pi/2 \\ \cos(2\delta)x_2(t^-) \end{pmatrix}, \quad x(t^-) \in S. \quad (9)$$

The Poincaré map  $\mathcal{P}$  of smooth system (8) induced by cross-section  $S$  can be computed in closed form employing the fact that the quantity

$$I(x_1, x_2) = \frac{1}{2}x_2^2 + \zeta \cos x_1 \quad (10)$$

stays constant on any solution  $x(t)$  of system (8). Combining  $\mathcal{P}$  with impact law (9), the formula for the full Poincaré map  $P$  reads as (see Byl and Tedrake 2009; Saglam et al. 2014)

$$P \begin{pmatrix} x_{1,began} \\ x_2 \end{pmatrix} = \begin{pmatrix} x_{1,begin} \\ \cos(2\delta) \sqrt{x_2^2 + 2\zeta \cos(x_{1,begin}) - 2\zeta \cos(x_{1,end})} \end{pmatrix},$$

provided that  $\delta \leq \pi$  and  $\zeta + \alpha - \pi/2 > 0$ . Another approach to construct a Poincaré map for Eqs. (8) and (9) is proposed in Coleman (2010).

**Proposition 1** *If*

$$\delta \geq \pi/2, \quad \zeta + \alpha - \pi/2 > 0,$$

*then conditions of Theorem 1 hold, and therefore, rimless wheel model (8) and (9) admits a unique orbitally stable limit cycle.*

A proof of Proposition 19 through a direct computation of the dynamics of  $P$  can be found, e.g., in Saglam et al. (2014).

Paper (Saglam et al. 2014) performed deterministic analysis of the case when the slope changes at each heel-strike and stays constant in between two successive heel-strikes. In addition, they (and Byl and Tedrake 2009) carried out a stochastic analysis of the case where the slope and the impact law change at each heel-strike independently (which can be viewed as a model

for a nonlinear terrain). A thorough analysis of the convergence of the dynamics of Eqs. (8) and (9) to the limit cycle along with the study of the respective basins of attraction is carried out in Coleman (2010). The work (Clark and Bloch 2020) extended rimless wheel model to the case of spokes of different lengths (called *nonuniform rimless wheel*), where a result of the type of a closing lemma on an interval (Young 1979) and the earlier achievement (Clark et al. 2020) were used, and equal angles between spokes (but the case when the angles between different spokes are different can be treated along the lines of this paper). Rimless wheel with slipping transitions is considered in Fazeli et al. (2017), Gamus and Or (2015), and Or (2014), where explicit computations are carried out. For the case of 3D rimless wheel, the Poincaré map turns out to be three-dimensional and no explicit formula for the Poincaré map is available. The paper (Coleman et al. 1997) proposed a perturbation approach (see section “Fixed Points of Perturbed Poincaré Maps”) to study fixed points of this map which

works when the slope of the ground is small. The paper (Coleman et al. 1997) discovered that in contrast with the rimless wheel in 2D, the 3D extension now features multiple walking cycles. The work (Smith and Berkemeier 1998) shows that the walking cycle is again unique in the case of two rigidly coupled 3D wheels (called *finite-width rimless wheel*).

### The Actuated Planar Biped (Poincaré Map Reduces to an Explicitly Computable Map)

In this section, the working of hybrid zero dynamics approach is reviewed by considering a three-link biped and following (Grizzle et al. 2001). The reader is referred to the book (Spong et al. 2020) for the background on the method of zero dynamics in smooth models of underactuated robots.

The equation of motion of three-link bipedal walker of Fig. 3 reads as

$$\begin{aligned}\ddot{\theta} &= F(\theta, \dot{\theta}) + G(\theta)u, \quad F: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3, \\ G(\theta) &\text{ is a } 3 \times 2 \text{ - matrix, } u \in \mathbb{R}^2,\end{aligned}\quad (11)$$

as long as leg  $\theta_1$  contacts the ground and leg  $\theta_2$  does not. The two-dimensional control  $u$  represents the two torques applied between the

torso and the stance leg, and between the torso and the swing leg. The control  $u$  will make sure that swing leg just slides along the floor (i.e., move at zero distance from the floor but does not penetrate the floor). Additional motors on the legs allow to push the swing leg just slightly out of the sagittal plane during the swing phase and to pull the leg back into the sagittal plane whenever an impact is desired. Specifically, the impact will be initiated when the angle of the stance leg attains a desired value,  $\theta_1^d$ . This leads to the following impact event

$$\begin{aligned}\left(\theta(t^+), \dot{\theta}(t^+)\right)^T &= \Delta \left( \left(\theta(t^-), \dot{\theta}(t^-)\right)^T \right), \\ \text{if } \theta_1(t^-) &= \theta_1^d.\end{aligned}\quad (12)$$

The control will also maintain the angle of the torso at some constant value  $\theta_3^d$ , while commanding the swing leg to behave as the mirror image of the stance leg (i.e.,  $\theta_2 = -\theta_1$ ). Therefore, the control  $u$  in Eq. (11) will satisfy the relations

$$\begin{aligned}y_1(t) &= \theta_3(t) - \theta_3^d \rightarrow 0 \quad \text{as } t \rightarrow \infty, \\ y_2(t) &= \theta_2(t) + \theta_1(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty.\end{aligned}\quad (13)$$

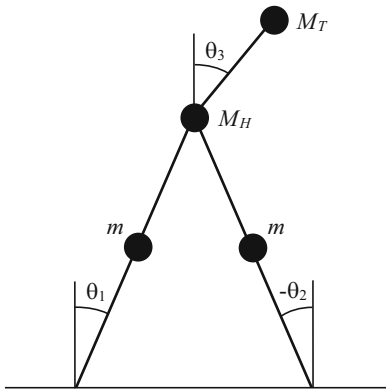
The Zero dynamics reduction is based on the change of variables  $(\theta_1, \theta_2, \theta_3) \mapsto (y_1, y_2, \theta_1)$  given by Eq. (13) which brings Eq. (11) to the form

$$\begin{aligned}\ddot{y} &= \zeta_0(y, \dot{y}_0, \theta_1, \dot{\theta}_1) + \zeta_1^u(y, \dot{y}_0, \theta_1, \dot{\theta}_1)u, \\ \ddot{\theta}_1 &= \zeta_0(y, \dot{y}_0, \theta_1, \dot{\theta}_1) + \zeta_1^u(y, \dot{y}_0, \theta_1, \dot{\theta}_1)u.\end{aligned}\quad (14)$$

According to Grizzle et al. (2001), the decoupling matrix  $\zeta_1^u(y, \dot{y}_0, \theta_1, \dot{\theta}_1)$  is invertible, if

$$-r(rM_H + rm + rM_T + lM_T \cos(\theta_1 - \theta_3)) \neq 0. \quad (15)$$

Therefore, if condition (15) holds, then introducing the new control variable



**Exact and Perturbation Methods in the Dynamics of Legged Locomotion, Fig. 3** The diagram of three-link planar bipedal robot and definitions of the angles  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  (see Grizzle et al. 2001, Fig. 1)

$$v = \zeta_0(y, \dot{y}, \theta_1, \dot{\theta}_1) + \zeta_1^u(y, \dot{y}, \theta_1, \dot{\theta}_1)u,$$

system (14) rewrites as

$$\ddot{y} = v, \ddot{\theta}_1 = \zeta_0(y, \dot{y}, \theta_1, \dot{\theta}_1) + \zeta_1(y, \dot{y}, \theta_1, \dot{\theta}_1)v. \quad (16)$$

Now one can use any control which stabilizes a double integrator to zero in finite time (e.g., Bhat and Bernstein (1998) or Santiesteban et al. (2010)) to drive the dynamics of the system of Eqs. (12) and (16) to

$$\begin{aligned} \ddot{\theta}_1 &= \zeta_0(0, 0, \theta_1, \dot{\theta}_1), \\ \theta_1(t^+) &= -\theta_1(t^-), & \text{if } \theta_1(t_0) = \theta_1^d, \\ \dot{\theta}_1(t^+) &= m\dot{\theta}_1(t^-) + m_0 \end{aligned} \quad (17)$$

see Grizzle et al. (2001, formulas (49), (50), and (52)). Equation (17) is called an *equation of hybrid zero dynamics*. According to Grizzle et al. (2001, formula (52)),

$$\zeta_0(0, 0, \theta_1, \dot{\theta}_1) = \bar{\zeta}_a(\theta_1) + \bar{\zeta}_b(\theta_1)\dot{\zeta}_1^2,$$

where

$$\begin{aligned} \bar{\zeta}_a(\theta_1) &= \frac{g}{r} \cdot \frac{(2m + M_T + M_H)r \sin(\theta_1) + M_T l \sin(\theta_3^d)}{mr + M_H r + M_T r + M_T l \cos(\theta_1 - \theta_3^d)}, \\ \bar{\zeta}_b(\theta_1) &= \frac{M_T l \sin(\theta_1 - \theta_3^d)}{mr + M_H r + M_T r + M_T l \cos(\theta_1 - \theta_3^d)}. \end{aligned}$$

In particular, Eq. (17) of zero dynamics for the biped of Fig. 3 reduce to the Eqs. (8) and (9) of rimless wheel model of Fig. 2 when  $M_T = 0$ . When  $M_T \neq 0$ , the second-order differential equation of (17) still admits an explicit integral of motion (Perram et al. 2003), and so the dynamics of Eq. (17) can be investigated along the lines of section “The Planar Rimless Wheel (Poincaré Map Is Explicitly Computable).”

The reader is referred to Freidovich et al. (2009), Grizzle et al. (2001, 2014), and Westervelt et al.

(2003, 2004) for further reading on hybrid zero dynamics approach. When the equation of hybrid zero dynamics is of higher dimension, it might not be possible to derive the Poincaré map in closed form. However, the method of *transversal linearization* (Manchester et al. 2011; Fevre et al. 2019; Shiriaev et al. 2010) allows to design a control that stabilizes the walking cycle of hybrid zero dynamics without computing the Poincaré map.

## Fixed Points of Perturbed Poincaré Maps

Consider a Poincaré map  $x \mapsto P(x, \delta)$  that depends on a parameter  $\delta$  and that admits a smooth one-parameter family of fixed points for  $\delta = 0$ , i.e.,  $P(\zeta(s), 0) = \zeta(s)$  for all  $s \in \mathbb{R}$ , with  $s \mapsto \zeta(s)$  being a  $C^1$  curve. Note, the latter property implies that  $P_x(\zeta(s), 0)\zeta'(s) = \zeta'(s)$ , where  $P_x(\zeta(s), 0)$  stays for the Jacobian. As a consequence, one of the eigenvalues of the matrix  $P_x(\zeta(s), 0)$  always equals 1 for all  $s \in \mathbb{R}$ .

Fix some  $s_0 \in \mathbb{R}$ , put

$$x_0 = \zeta(s_0),$$

and denote by  $\rho_1, \dots, \rho_m$  the eigenvalues of  $P_x(x_0, 0)$  with  $\rho_1$  being equal 1: This notation assumes that the algebraic multiplicity of each eigenvalue is 1. Assume further that

$$|\rho_2| \neq 1, \dots, |\rho_m| \neq 1. \quad (18)$$

Assumption (18) allows to denote by  $y$  and  $z$  the eigenvectors of  $P_x(x_0, \omega_0, 0)$  and  $P_x(x_0, \omega_0, 0)^T$  that correspond to the eigenvalue 1 and satisfy

$$z^T y > 0. \quad (19)$$

The ideas of the following theorem are due to Malkin (1956, Chap. VI), Glover et al. (1989), and Loud (1959) (see also Kamenskii et al. 2011; Kumar et al. 2022; Makarenkov 2020; Makarenkov and Ortega 2011 for more refined formulations and proofs).

**Theorem 2** *Let  $P$  be a  $C^3$  function and assume that Eq. (18) holds. If, for each  $\delta \in \mathbb{R}$ , the*

Poincare map  $x \mapsto P(x, \delta)$  admits a fixed point  $x_\delta$  such that

$$\|x_\delta - x_0\| \leq N\delta, \quad (20)$$

for some  $N > 0$ , and for all  $|\delta|$  sufficiently small, then

$$z^T P_\delta(x_0, 0) = 0. \quad (21)$$

Assume that the eigenvector  $z$  of  $P_x(\xi(s_0), 0)^T$  that corresponds to the eigenvalue 1 does not depend on  $s_0$ , i.e.,

$$z^T (P_x(\xi(s), 0) - I) = 0, \quad \text{for all } s \in \mathbb{R}. \quad (22)$$

If, in addition to Eqs. (21) and (22), it holds that

$$z^T P_{\delta x}(x_0, 0)y \neq 0, \quad (23)$$

then, for all  $|\delta|$  sufficiently small, the Poincare map  $x \mapsto P(x, \delta)$  does indeed have a fixed point  $x_\delta$  that satisfies Eq. (20) for some  $N > 0$ . The fixed point  $(\theta_\delta, \omega_\delta)$  is asymptotically stable, if Eq. (18) holds in the stronger sense

$$|\rho_2| < 1, \dots, |\rho_m| < 1, \quad (24)$$

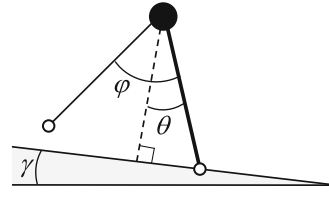
and if Eq. (23) holds in the stronger sense

$$z^T P_{\delta x}(x_0, 0)y < 0. \quad (25)$$

The next section follows (Garcia et al. 1998; Makarenkov 2020) to discuss the best (to the author's knowledge) application of Theorem 2 discovered by Garcia, Chatterjee, Ruina, and Coleman (1998).

### Compass-Gait Biped (Poincare Map Is a Perturbation of an Explicitly Computable Map)

In his celebrated paper (McGeer 1990), McGeer proposed to view the passive bipedal walker of Fig. 4 as a double pendulum, where the roles of



**Exact and Perturbation Methods in the Dynamics of Legged Locomotion, Fig. 4** The double pendulum model of a planar biped (see Garcia et al. 1998, Fig. 1)

the two links swap each time the heelstrike occurs (i.e., swing leg collides with the ground). Assuming that the masses of the feet are negligible compared to the weight of the body, the paper (Garcia et al. 1998) formulated double pendulum model of bipedal walker between heelstrikes as

$$\begin{aligned} \ddot{\theta} - \sin(\theta - \gamma) &= 0, \\ \ddot{\theta} - \ddot{\phi} + \dot{\theta}^2 \sin \phi - \cos(\theta - \gamma) \sin \phi &= 0, \end{aligned} \quad (26)$$

where  $\gamma \geq 0$  is the slope of the ground (see Fig. 4). The length of the legs and the gravity acceleration constants have been removed from Eq. (26) by a suitable time rescaling (see Garcia et al. 1998).

When a heelstrike occurs (i.e., when  $\phi = 2\theta$ ), the stance and swing legs swap their roles and the state vector  $(\theta, \dot{\theta}, \phi, \dot{\phi})^T$  jumps as

$$\begin{aligned} &(\theta(t), \dot{\theta}(t), \phi(t), \dot{\phi}(t))^T \\ &= \Delta \left( (\theta(t^-), \dot{\theta}(t^-))^T \right), \quad \text{if } \phi(t) = 2\theta(t). \end{aligned} \quad (27)$$

The model (26) and (27) is a particular case of a compass-gait biped. Using Newton's method, McGeer found that a certain linearization of switched system (26) and (27) admits a limit cycle, whose period is close to  $T = 3.8$  for small values of slope  $\gamma > 0$ .

A justification of the existence of such a limit cycle was offered in Garcia et al. (1998), where the change of the variables

$$\gamma = \delta^{3/2}, \quad \theta(t) = \delta^{1/2}\Theta(t), \quad \phi(t) = \delta^{1/2}\Phi(t) \quad (28)$$

is proposed to expand Eqs. (26) and (27) in the powers of small parameter  $\delta > 0$  and to investigate the existence of a limit cycle based on perturbation theory.

Incorporating the change of the variables (28) into the switched system (26) and (27) and using that

$$\begin{aligned} \sin \tau &= \tau - \frac{\tau^3}{3!} + \frac{\tau^5}{5!} - \frac{\tau^7}{7!} + \dots, \\ \cos \tau &= 1 - \frac{\tau^2}{2!} + \frac{\tau^4}{4!} - \frac{\tau^6}{6!} + \dots, \end{aligned}$$

the authors of Garcia et al. (1998) obtained

$$\begin{aligned} \ddot{\Theta} - (\Theta - \delta) + \frac{1}{6}\delta\Theta^3 + o_1(\delta) &= 0, \\ \ddot{\Phi} - \Phi - \ddot{\Phi} + \delta\dot{\Theta}^2\Phi + \frac{1}{2}\delta\Theta^2\dot{\Phi} + \frac{1}{6}\delta\Phi^3 + o_2(\delta) &= 0, \end{aligned} \quad (29)$$

$$\begin{aligned} &(\Theta(t), \dot{\Theta}(t), \Phi(t), \dot{\Phi}(t))^T \\ &= \Delta\left((\Theta(t^-), \dot{\Theta}(t^-))^T, \delta\right), \text{ if } \Phi(t) \\ &= 2\Theta(t), \end{aligned} \quad (30)$$

where  $o_i(\delta)$  stays for the remainders (dependent on  $\Theta$  and  $\Phi$ ) such that  $o_i(\delta)/\delta \rightarrow 0$  as  $\delta \rightarrow 0$  uniformly with respect to  $(\Theta, \Phi)$  from any compact set. The reader is referred to Yudaev et al. (2017) for asymptotic expansion (29) for the case where the role of the small parameter is played by both the slope of the ground and the ratio of the masses of the feet and the mass of body.

The fact that the impact law in Eq. (30) depends on only two-phase variables allows to introduce a two-dimensional Poincaré map with small parameter  $\delta$ . Indeed, since, according to Garcia et al. (1998), the image  $\Delta\left((\Theta(t^-), \dot{\Theta}(t^-))^T, \delta\right)$  of the impact is of the form

$$\begin{aligned} &\Delta\left((\Theta(t^-), \dot{\Theta}(t^-))^T, \delta\right) \\ &= (\theta, \omega, 2\theta, 2\delta\theta^2\omega + o_4(\delta)\omega)^T, \end{aligned}$$

one only needs the solution  $t \mapsto (\Theta, \dot{\Theta}, \Phi, \dot{\Phi})(t, \theta, \omega, \delta)$  of Eq. (29) that originates from the initial condition

$$\begin{aligned} &(\Theta(0), \dot{\Theta}(0), \Phi(0), \dot{\Phi}(0)) \\ &= (\theta, \omega, 2\theta, 2\delta\theta^2\omega + o_4(\delta)\omega). \end{aligned} \quad (31)$$

The required Poincaré map then reads as

$$P(\theta, \omega, \delta) = J((\Theta, \dot{\Theta})(T(\theta, \omega, \delta), \theta, \omega, \delta)\delta) \quad (32)$$

where

$$J(\Theta, \Omega, \delta) = \begin{pmatrix} -1 & 0 \\ 0 & 1 - 2\delta\Theta^2 + o_3(\delta) \end{pmatrix} \begin{pmatrix} \Theta \\ \Omega \end{pmatrix}, \quad (33)$$

and where  $T(\theta, \omega, \delta)$  (the time of nearest heel-strike) is the time satisfying

$$\begin{aligned} &\Phi(T(\theta, \omega, \delta), \theta, \omega, \delta) = 2\Theta(T(\theta, \omega, \delta), \theta, \omega, \delta), \\ &\Phi(t, \theta, \omega, \delta) \neq 2\Theta(t, \theta, \omega, \delta), \quad t \in (0, T(\theta, \omega, \delta)). \end{aligned} \quad (34)$$

Note, following Garcia et al. (1998), the Poincaré map  $P$  is introduced based on Definition 7; see Remark.

Using explicit solvability of Eq. (29) for  $\delta = 0$ , the authors of Garcia et al. (1998) conclude that the Poincaré map  $P$  admits the following family of fixed points

$$\omega = \alpha(T)\theta, \quad (35)$$

where

$$\begin{aligned} \alpha(T) &= -\frac{1 + e^T}{-1 + e^T}, \quad -3 + 3e^T \\ &\quad + 3(-1 + e^T)\cos T + \sin T + e^T \sin T \\ &= 0. \end{aligned} \quad (36)$$

For every root of Eq. (36), formula (35) defines a family of fixed points of  $(\theta, \omega) \mapsto P(\theta, \omega, 0)$ . The survey reviews the available results for the root  $T_2 = 3.81209$  called the *long period*.



Plugging  $\omega = \alpha(T_2)\theta$  into the impact condition  $\Phi(t, \theta, \omega, 0) = 2\Theta(t, \theta, \omega, 0)$  gives approximately

$$\begin{aligned} & -1.5339e^{-t} + 0.0339021e^t + 1.5 \cos t \\ & + 0.522601 \sin t = 0, \end{aligned} \quad (39)$$

whose only solution on  $(0, T_2)$  is  $T_2/2$  where one has

$$\Theta(T_2/2, \theta, \omega, 0) = \Phi(T_2/2, \theta, \omega, 0) = 0. \quad (37)$$

The property (37) corresponds to the event where the two legs coincide. Though Eq. (37) formally implies a heel-strike, it corresponds to just grazing of the swing leg through the floor and no impact event physically occurs. If the value of  $\gamma$  increases, then, formally speaking, an impact near  $T = T_2/2$  may occur, but it is common to ignore such a near-grazing impact as motivated by the experiments (the experimental passive walker makes slight swings in the frontal direction which rules out the impact at  $T = T_2/2$ ; see Cox 2009).

The reader can consult (Makarenkov 2020) for the computation of  $P_{(\theta, \omega)}(\theta, \alpha(T_2)\theta, 0)$  and  $P_\delta(\theta, \alpha(T_2)\theta, 0)$  which turn out to be

$$P_{(\theta, \omega)}(\theta, \alpha(T_2)\theta, 0) = \begin{pmatrix} -5.07075 & -5.8082 \\ 5.8082 & 6.55701 \end{pmatrix}, \quad (38)$$

$$P_\delta(\theta, \alpha(T_2)\theta, 0) = \begin{pmatrix} 5.85426 + 0.348762\theta^3 \\ -7.51458 + 1.75673\theta^3 \end{pmatrix}.$$

The eigenvalues of Eq. (38) are 1 and

$$\rho = 0.48626, \quad (40)$$

so that condition (24) holds. Computing an eigenvector  $z$  of the transpose of matrix (38) for the eigenvalue 1, one gets

$$z = (-0.69131, -0.722559)^T.$$

Therefore, taking into account the relation (35) between  $\theta_0$  and  $\omega_0$ , the necessary condition (21) of Theorem 2 takes the form

$$1.38262 - 1.51044(\theta_0)^3 = 0.$$

The solution of this equation is (Garcia et al. 1998; Makarenkov 2020)

$$\theta_0 = 0.970956. \quad (41)$$

To verify sufficiency and stability condition of Theorem 2, the quantity  $P_{\delta(\theta, \omega)}(\theta, \alpha(T_2)\theta, 0)$  is required, which computes as (Makarenkov 2020)

$$P_{\delta(\theta, \omega)}(\theta, \alpha(T_2)\theta, 0) = \frac{1}{\theta} \begin{pmatrix} 218.645 - 3.99563\theta^3 & 209.189 - 4.82387\theta^3 \\ -218.652 - 30.1237\theta^3 & -209.195 - 33.8632\theta^3 \end{pmatrix}. \quad (42)$$

To check the stability condition (25), it remains to compute the eigenvector  $y$  of matrix (38) which corresponds to the eigenvalue 1 and satisfies the sign condition (19) with the vector  $z$  obtained above. Such a computation leads to

$$y = (0.69131, -0.722559)^T.$$

Using formula (42) and the value  $\theta_0$  given by Eq. (41), one gets

$$\lambda_* = z^T P_{\delta(\theta, \omega)}(\theta_0, \alpha(T_2)\theta_0, 0)y = -2.95323, \quad (43)$$

so that both the conditions (23) and (25) hold.

Using Theorem 2, the following proposition can be concluded.

**Proposition 2** *For all  $\delta > 0$  sufficiently small, the impact system (29) and (30) admits an asymptotically stable limit cycle with the initial condition  $(\theta_\delta, \omega_\delta, 2\theta_\delta, (2\delta(\theta_\delta)^2 + o_4(\delta)\omega_\delta))$ , where  $(\theta_\delta, \omega_\delta) \rightarrow (\theta_0, \alpha(T_2)\theta_0)$  as  $\delta \rightarrow 0$ . The cycle experiences exactly one impact per the period  $T_\delta > 0$  and  $T_\delta \rightarrow T_2$  as  $\delta \rightarrow 0$ . Accordingly, the initial model (26) and (27) admits a walking cycle obtained from the limit cycle of Eqs. (29) and (30) over the change of the variables (28).*

**Remark 2** Note, the origin is a possible solution to Eqs. (26) and (27) when  $\gamma = 0$  (one can look up the formula for  $\Delta$  in Garcia et al. (1998) to verify that  $\Delta$  maps zero to zero). And the origin belongs to the switching manifold defined in Eq. (27). On the other hand, the limit cycle  $t \mapsto (\theta_\gamma(t), \dot{\theta}_\gamma(t), \phi_\gamma(t), \dot{\phi}_\gamma(t))$  of Eqs. (26) and (27) given by Corollary 2 encounters one impact per period and converges to the origin as  $\gamma \rightarrow 0$ . Therefore, the occurrence of the cycle  $t \mapsto (\theta_\gamma(t), \dot{\theta}_\gamma(t), \phi_\gamma(t), \dot{\phi}_\gamma(t))$  from the origin as  $\gamma > 0$  crosses zero is a border-collision bifurcation according to the terminology of nonsmooth dynamical systems (see textbooks (di Bernardo et al. 2008a; Simpson 2010)). In other words, Garcia et al. (1998) used a blow-up technique in order to establish border-collision bifurcation of limit cycles in a four-dimensional system before the relevant results (di Bernardo et al. 2008b; Makarenkov and Niwanthi Wadippuli Achchige 2018; Simpson and Meiss 2007; Zou et al. 2006) appeared in the literature on nonsmooth dynamical systems. A control strategy that uses the existence of a walking cycle in Eqs. (26) and (27) for shallow slopes in order to establish the existence and stability

of a walking gait on bigger slopes is proposed in Hu et al. (2011).

Control strategies to improve stability of limit cycles of Eqs. (26) and (27) are developed in Asano and Luo (2009), Byl and Tedrake (2008), Chyou et al. (2011), Crews and Travers (2020), Holm and Spong (2008), Iida and Tedrake (2010), Sidorov and Zacksenhouse (2019), Spong (1999), Safa et al. (2016), and Wisse et al. (2005). In particular, enlarging the domain of attraction of the compass-gait biped is addressed in Iida and Tedrake (2010), Sidorov and Zacksenhouse (2019), and Wisse et al. (2005). Passive walking over slippery terrain is addressed in Chen and Goodwine (2021), Gamus and Or (2015), and Or (2014). The case of varying or rough terrain is considered in Bhounsule and Zamani (2017), Crews and Travers (2020), Efimov et al. (2014), Iida and Tedrake (2010), and Liu et al. (2020).

## Future Directions

This section lists opportunities for perturbation theory in the field of legged locomotion. The universal goal throughout the list is to understand why the planar passive biped is much more stable in the experiment compared to its mathematical model.

### Nonperiodic Motions (Existence of Invariant Tori)

Stability of robotic walking to changing terrain or gait primitive is addressed in Burridge et al. (1999), Byl and Tedrake (2009), Efimov et al. (2014), Gregg et al. (2012), Manchester et al. (2011), Veer et al. (2017), Saglam et al. (2014), Shiriaev et al. (2010), Sovero et al. (2015), Tang et al. (2018), and Veer and Poulakakis (2020). These papers either use Lyapunov funneling, jump-Riccati equation, or stochastic framework to design the stabilizing control. Furthermore, the notions of stability appropriate for nonperiodic motions are proposed in Byl and Tedrake (2008), Efimov et al. (2014), Su and Dingwell (2007), and Yang et al.

(2009). The abovementioned papers focus on control design. Can stable walking in changing (state-dependent or time-dependent) environment be obtained as just a perturbation of a limit cycle walking along a constant slope? This fundamental question about structural stability of walking cycle to nonautonomous perturbations (answered for smooth differential equations in Levinson's (1950)) is not understood for impact systems yet.

In (1950), Levinson discovered that, for smooth systems, small perturbation of a system with an attracting limit cycle by an arbitrary time-dependent force leads to a system with an attracting torus (and nothing, in general, can be said about the dynamics on the torus). While there have been several important results on persistence of invariant manifolds for nonsmooth dynamical systems (Szalai and Osinga 2008a; Weiss et al. 2012), no extensions of the Levinson's result for systems with impacts are currently available. Obtaining such an extension will not only shed light on the presence of attracting aperiodic gaits in robotic locomotion, but also contribute to the problem of walking on uneven terrain and maneuvering.

Whereas a nonsmooth analogue of Levinson's theorem can be applied to investigate nonperiodic motions in self-oscillating impact systems obtained via hybrid zero dynamics approach, other methods are required for self-oscillating impact systems obtained via the perturbation approach, which contain a small parameter already (and arguing via the Levinson's approach would require a second small parameter of higher order making the model less realistic). In particular, the analysis of nonperiodic motions of robotic walkers will greatly benefit from the extension of the perturbation approach of section "Fixed Points of Perturbed Poincaré Maps" to the case when the perturbation term of the right-hand-side of system (29) and (30) depends on time in a nonperiodic way. The corresponding method is referred to the names of Melnikov (Guckenheimer and Holmes 1990; Melnikov 1963), Bogoliubov and Mitropolsky (1961, Sec. 29), Malkin (1956), and Loud (1959) in the classical smooth literature. Perturbation methods for impact systems where the impact law

depends on time did receive some attention (Filatov 2008; Graef et al. 2013; Iannelli et al. 2006; Perestyuk et al. 2011; Thomsen and Fidler 2008) (i.e., when the 'if' condition of law (27) is expressed in terms of time); however, only partial results are available in the case of state-dependent laws of impact; see Battelli and Feckan (2013a) and Sfecci (2017) for existence results and Newman and Makarenkov (2015) for stability results in the case of 1:1 resonance. Following the ideas of Bogoliubov and Mitropolsky (1961, Sec. 29), dropping the assumption of periodicity of time-dependent perturbation will require the existence of some average of the perturbation term over time.

### Varying Impact Law (Dependence of Stable Manifolds on Parameters)

The authors of Iida and Tedrake (2010), Sidorov and Zacksenhouse (2019), and Wisse et al. (2005) succeeded to enlarge the domain of attraction of the limit cycle of the compass-gait biped by amending the impact law (27). The paper (Garcia et al. 2000) discussed the effect of the impact law on the efficiency of bipedal locomotion (while also considering various distributions of masses in the walker). The work (Safa et al. 2016) amended the surface in order to control the eigenvalues of the Jacobian  $P_{(\theta, \omega)}(\theta_\delta, \omega_\delta, \delta)$  (Safa et al. (2016) used several different models for the Jacobian  $P$ ). At the same time, the fact that the eigenvectors of the Jacobian (38) of the reduced (i.e., unperturbed) Poincaré map of Eqs. (26) and (27) are almost parallel has been overlooked in the literature so far. Whether or not the amendment of the angle between the eigenvectors of Eq. (38) is the underlying principle of the control strategy of Sidorov and Zacksenhouse (2019) is not clarified in the literature yet. Using the ideas of Sidorov and Zacksenhouse (2019) to develop a tool that controls the angle between the eigenvectors of Eq. (38) (and, more globally, between the stable manifolds of walking limit cycle) by changing the impact law (27) and by even allowing some sliding after impact (see Gamus and Or 2015) can pave a way to analytic results about the enlargement of the domain of attraction of the walking limit cycle of compass-gait biped.

### The Effect of Leg Scuffing on Stability of the Walking Cycle (Grazing Bifurcations)

Foot scuffing of compass-gait biped (almost tangential collision of the swing leg with the ground at the time when positions of the two legs coincide) is usually explicitly ignored in the literature because the experimental biped seems to swing in frontal direction slightly at the moment of expected scuffing so that the swing leg actually misses to collide with the ground. Accounting for leg scuffing would require investigating stability of walking cycles which are almost tangent to the ground. Mathematically speaking, one would need to spot a small bifurcation parameter whose smaller values yield more tangential collisions of the walking cycle with the ground. The occurrence (under varying parameter) of limit cycles which are almost tangential to the impact surface from a cycle which is fully tangential to the impact surface is called *grazing bifurcation*. Though grazing bifurcation in impact systems destroys stability generically (Budd and Piiroinen 2006; Kryzhevich and Wiercigroch 2012; Nordmark 1991), it may help to stabilize the system when certain degeneracy assumption holds, as, e.g., in atomic force microscopy applications (Misra et al. 2010; Yagasaki 2004) (see also Rom-Kedar and Turaev (1999) and Turaev and Rom-Kedar (1998) for a relevant phenomenon in billiards dynamics). And there is a good chance that the required degeneracy condition for the passive biped holds when the role of the abovementioned bifurcation parameter is played by the slope of the ground (i.e.,  $\delta$  of section “Compass-Gait Bipod (Poincare Map Is a Perturbation of an Explicitly Computable Map)”). Indeed, the impact system (29) and (30) admits an entire family of cycles and is, therefore, degenerate for  $\delta = 0$ . That is why accounting for scuffing could give an explanation for the remarkable stability of the experimental passive walker. Referring to the particular equations of section “Compass-Gait Bipod (Poincare Map Is a Perturbation of an Explicitly Computable Map),” the problem consists in accounting for possible near-grazing impacts around  $t = T_2/2$ ; see Eq. (37).

An alternative framework for the analysis of stabilizing effects associated with near-grazing

impacts has been developed in Chillingworth (2010), which discovered that the points of collision with the ground form invariant manifolds bounded by a so-called *discontinuity arc*. Understanding the dependence of these invariant manifolds and the discontinuity arc on the perturbation parameter  $\delta$  of the passive walker (the slope of the ground, see section “Compass-Gait Bipod (Poincare Map Is a Perturbation of an Explicitly Computable Map)”) is an open problem, whose resolution will provide a new tool for the analysis of the stability of walking cycles with near-grazing impacts. A step in this direction is accomplished in Kryzhevich and Wiercigroch (2012).

The existence of the phases where the swing leg slides along the ground during scuffing is possible in the presence of dry friction. The occurrence of limit cycles with small sliding phases from a limit cycle without sliding is referred to as *grazing-sliding bifurcation*. This bifurcation is extensively studied in the literature (Glendinning et al. 2012, 2016; Glendinning and Jeffrey 2015; Li and Chen 2020; Makarenkov 2017b; Nordmark and Kowalczyk 2006; Simpson 2017; Szalai and Osinga 2008b, 2009) for systems with discontinuous right-hand-sides (where the discontinuity can be used to model dry friction), but no application to robotic scuffing was considered. Allowing for dry friction and investigating possible sliding of swing leg during scuffing is another opportunity which can contribute to understanding stability of the passive walker. Indeed, the stabilizing effect of near-tangent collisions has been already observed in the context of, e.g., power converters (see Benmiloud et al. 2019).

Developing a tool that analyzes stability of bipedal walker with sliding (both at the time of impact as in section “Varying Impact Law (Dependence of Stable Manifolds on Parameters)” or at the time of scuffing as in section “The Effect of Leg Scuffing on Stability of the Walking Cycle (Grazing Bifurcations)”) might be of outstanding value for the efficiency of multi-legged locomotion. Indeed, cockroaches’ legs slip for 15–18% of the time (see Zhao and Revzen 2020), and the work (Gamus and Or 2015) documented at least double energy gain when

allowing sliding in the model of bipedal walker. The interested reader is advised to consult (Gamus et al. 2020, 2021) (direct computations and experiments) and (Moreau sweeping process approach) (Gidoni 2018; Gidoni and DeSimone 2017) for the recent progress and further motivations toward studying slippery phases of multilegged robots.

### The Effect of Soft Ground on Stability of the Walking Cycle (Singular Perturbations)

The dynamics of collisions with soft impacts (i.e., slightly inelastic as opposed to absolutely elastic) has been investigated, e.g., in Ivanov (1994, 1996). These studies can be used to incorporate soft ground into the zero dynamics-based analysis of the actuated biped of section “[The Actuated Planar Biped \(Poincaré Map Reduces to an Explicitly Computable Map\)](#).” A commonly used mathematical approach which allows to replace rigid ground by an array of (strong) springs is a so-called *theory of singular perturbations* (Jones 1995; Verhulst 2005), where a small parameter in front of the derivatives of some of the differential equations corresponds to the reciprocal of springs’ stiffnesses. In the singular perturbations approach, a small part of the limit cycles is governed by a separate system of differential equations that models the dynamics of the ground. Therefore, two small parameters would occur when a singular perturbation approach is used to model the ground, and a regular perturbation approach is used to model the passive biped (as it is done in section “[Compass-Gait Biped \(Poincaré Map Is a Perturbation of an Explicitly Computable Map\)](#)”). There has been significant progress lately (Battelli and Feckan 2013b; Makarenkov and Verhulst 2021) toward the development of a theory which is capable to analyze models where singular and regular perturbations are present simultaneously. There is a good chance the methods of Battelli and Feckan (2013b) and Makarenkov and Verhulst (2021) will allow to understand the conditions under which soft ground enlarges the domain of attraction of the walking cycle and explain stability of the experimental passive walker.

### The Effect of Accounting for the Third Dimension (Higher-Dimensional Perturbation Analysis)

Paper (Sabaapour et al. 2015) reports that 3D motions of the 3D rimless wheel of Coleman et al. (1997) are more stable compared to 2D motions of the same wheel (the 3D rimless wheel admits different types of motion). Therefore, the frontal dimension of the experimental compass-gait biped may play a crucial role in the stability of the walking cycle. A model where a planar passive biped is gradually transformed into a biped in dimension 3 under a varying parameter is proposed in Adolfsson et al. (2001) (see also Piironen et al. 2003). The achievement of Adolfsson et al. (2001) creates fruitful opportunities for the use of perturbation theory in order to examine the dependence of stability properties of walking cycle when the varying parameter of Adolfsson et al. (2001) adds the third dimension. In particular, Adolfsson et al. (2001) provides a tool (for future researchers) to understand how the dimension adding parameter influences the angle between the eigenvectors of the Jacobian of the Poincaré map (see section “[Varying Impact Law \(Dependence of Stable Manifolds on Parameters\)](#)”). To implement this strategy fully analytically, one would require introducing a small parameter that measures the slope of the ground (as in the analysis of the passive biped of section “[Compass-Gait Biped \(Poincaré Map Is a Perturbation of an Explicitly Computable Map\)](#)”) and following the lines of the perturbation approach of section “[Compass-Gait Biped \(Poincaré Map Is a Perturbation of an Explicitly Computable Map\)](#).” Of course, computation of the quantities required for the application of Theorem 2 becomes more complicate, but one can use the quantities already computed for the 2D model in section “[Compass-Gait Biped \(Poincaré Map Is a Perturbation of an Explicitly Computable Map\)](#)” and then continue them to the 3D model using the approach of Adolfsson et al. (2001) (as was done to, e.g., establish that  $\det(A_2) > 0$  in the proof of the Theorem of Zhang et al. (2016)). An ambitious goal would be to compute the quantities of Theorem 2 for an arbitrary 3D passive biped with small



slope of ground (not only those close to the 2D limit). Accomplishing this goal will allow engineers to use 3D passive biped as the equation of zero dynamics (17), thus increasing the efficiency of locomotion even further. Moreover, having an analytic perturbation tool for 3D passive walkers will allow to clarify the stabilizing effect of robotic arms. Inspirational experiments in this directions are carried out in Collins et al. (2009), de Graaf et al. (2019), and Sobajima et al. (2013).

On top of Adolfsson et al. (2001), 3D bipeds that walk under the action of just gravity are considered in Collins et al. (2001), Pankov (2021), and Tedrake et al. (2004). Efficiently controlled walking of 3D bipeds is addressed in Adolfsson et al. (2001), Beigzadeh et al. (2018), Ikeda and Toyama (2015), Spong et al. (2007), Spong and Bullo (2005), Tedrake et al. (2004), Wang et al. (2009), and Wisse et al. (2001). The models of these papers might be of further help with the choice of small parameters that makes application of perturbation theory most simple.

**Acknowledgments** The study was funded by the National Science Foundation (grant CMMI-1916876).

## Bibliography

- Adolfsson J, Dankowicz H, Nordmark A (2001) 3D passive walkers: finding periodic gaits in the presence of discontinuities. *Nonlinear Dyn* 24:205–229
- Andronov AA, Vitt AA, Khaikin SE (1987) *Theory of oscillators*. Russian edition: Immirzi F (1987). Reprint of the 1966 translation. Dover Publications, Inc., New York, xxxiv+815
- Asano F (2015) Fully analytical solution to discrete behavior of hybrid zero dynamics in limit cycle walking with constraint on impact posture. *Multibody Syst Dyn* 35(2):191–213
- Asano F, Luo ZW (2009) Asymptotically stable biped gait generation based on stability principle of rimless wheel. *Robotica* 27(6):949–958
- Babitsky VI (1998) *Theory of vibro-impact systems and applications*, in: foundations of engineering mechanics. Springer, Berlin
- Battelli F, Feckan M (2013a) Chaos in forced impact systems. *Discrete Contin Dyn Syst Ser S* 6(4):861–890
- Battelli F, Feckan M (2013b) Fast-slow dynamical approximation of forced impact systems near periodic solutions. *Bound Value Probl* 2013:71, 33 pages
- Beigzadeh B, Sabaapour MR, Yazdi MRH, Raahemifar K (2018) From a 3D passive biped walker to a 3D passivity-based controlled robot. *Int J Humanoid Robot* 15:1850009, 27 pages
- Benmiloud M, Benalia A, Djemai M, Defoort M (2019) On the local stabilization of hybrid limit cycles in switched affine systems. *IEEE Trans Autom Control* 64(2): 841–846
- Bhat SP, Bernstein DS (1998) Continuous finite-time stabilization of the translational and rotational double integrators. *IEEE Trans Autom Control* 43(5):678–682
- Bhounsule PA, Zamani A (2017) Stable bipedal walking with a swing-leg protraction strategy. *J Biomech* 51: 123–127
- Bogoliubov NN, Mitropolsky YA (1961) *Asymptotic methods in the theory of non-linear oscillations*, International monographs on advanced mathematics and physics. Hindustan Publishing Corp./Gordon and Breach Science Publishers, Delhi/New York. Translated from the second revised Russian edition
- Budd CJ, Piironen PT (2006) Corner bifurcations in non-smoothly forced impact oscillators. *Phys D* 220(2): 127–145
- Burridge R, Rizzi A, Koditschek D (1999) Sequential composition of dynamically dexterous robot behaviors. *Int J Robot Res* 18(6):534–555
- Byl K, Tedrake R (2008) Approximate optimal control of the compass gait on rough terrain. In: *Proc 2008 IEEE international conference on robotics and automation (ICRA)*, Pasadena, vols 1–9, pp 1258–1263
- Byl K, Tedrake R (2009) Metastable walking machines. *Int J Robot Res* 28:1040–1064
- Castillo J (2020) The pseudo-Hopf bifurcation and derived attractors in 3D Filippov linear systems with a Teixeira singularity. *Chaos* 30(11):113101, 14 pages
- Chen T, Goodwine B (2021) Robust gait design for a compass gait biped on slippery surfaces. *Robot Auton Syst* 140:103762
- Chillingworth DRJ (2010) Dynamics of an impact oscillator near a degenerate graze. *Nonlinearity* 23(11):2723–2748
- Chyou T, Liddell GF, Paulin MG (2011) An upper-body can improve the stability and efficiency of passive dynamic walking. *J Theor Biol* 285(1):126–135
- Clark W, Bloch A (2020) A Poincaré-Bendixson theorem for hybrid dynamical systems on directed graphs. *Math Control Signals Syst* 32(1):1–18
- Clark W, Bloch A, Colombo L (2020) A Poincaré-Bendixson theorem for hybrid systems. *Math Control Relat Fields* 10(1):27–45
- Coleman MJ (2010) Dynamics and stability of a rimless spoked wheel: a simple 2D system with impacts. *Dyn Syst* 25(2):215–238
- Coleman MJ, Chatterjee A, Ruina A (1997) Motions of a rimless spoked wheel: a simple three-dimensional system with impacts. *Dyn Stab Syst* 12(3):139–159
- Collins SH, Wisse M, Ruina A (2001) A 3-D passive dynamic walking robot with two legs and knees. *Int J Robot Res* 20(7):607–615



- Collins S, Ruina A, Tedrake R, Wisse M (2005) Efficient bipedal robots based on passive-dynamic walkers. *Science* 307(5712):1082–1085
- Collins SH, Adamczyk PG, Kuo AD (2009) Dynamic arm swinging in human walking. *Proc R Soc B* 276: 3679–3688
- Cox B (2009) <https://www.youtube.com/watch?v=N64KOQkbyil>
- Crews S, Travers M (2020) Energy management through footstep selection for bipedal robots. *IEEE Robot Autom Lett* 5(4):5485–5493
- Cristiano R, Pagano DJ, Carvalho T, Tonon DJ (2019) Bifurcations at a degenerate two-fold singularity and crossing limit cycles. *J Differ Equ* 268(1):115140
- de Graaf ML, Hubert J, Houdijk H, Bruijn SM (2019) Influence of arm swing on cost of transport during walking. *Biol Open* 8(6):bio039263
- DeCarlo RA, Branicky MS, Pettersson S, Lennartson B (2000) Perspectives and results on the stability and stabilizability of hybrid systems. *Proc IEEE* 88(7): 1069–1082
- di Bernardo M, Budd CJ, Champneys AR, Kowalczyk P (2008a) Piecewise-smooth dynamical systems. Theory and applications, Applied mathematical sciences, vol 163. Springer, London, p xxii+481
- di Bernardo M, Nordmark A, Olivar G (2008b) Discontinuity-induced bifurcations of equilibria in piecewise-smooth and impacting dynamical systems. *Phys D Nonlinear Phenom* 237:119–136
- Edmond JF, Thibault L (2006) BV solutions of nonconvex sweeping process differential inclusion with perturbation. *J Differ Equ* 226(1):135–179
- Efimov D, Perruquetti W, Shiriaev A (2014) On existence of oscillations in hybrid systems. *Nonlinear Anal Hybrid Syst* 12:104–116
- Fazeli N, Kolbert R, Tedrake R, Rodriguez A (2017) Parameter and contact force estimation of planar rigid-bodies undergoing frictional contact. *Int J Robot Res* 36(13–14):1437–1454
- Fevre M, Goodwine B, Schmiedeler JP (2019) *Int J Robot Res* 38(10–11):1307–1323
- Filatov OP (2008) Averaging of systems of differential inclusions with slow and fast variables. *Diff Uravn* 44(3):336–349, 429–430; translation in *Differ Equ* 44(3):349–363 (2008)
- Freidovich LB, Mettin U, Shiriaev AS, Spong MW (2009) A passive 2-DOF walker: hunting for gaits using virtual holonomic constraints. *IEEE Trans Robot* 25(5):1202–1208
- Gamus B, Or Y (2015) Dynamic bipedal walking under stick-slip transitions. *SIAM J Appl Dyn Syst* 14(2): 609–642
- Gamus B, Salem L, Gat AD, Or Y (2020) Understanding inchworm crawling for soft-robotics. *IEEE Robot Autom Lett* 5(2):1397–1404
- Gamus B, Gat AD, Or Y (2021) Dynamic inchworm crawling: performance analysis and optimization of a three-link robot. *IEEE Robot Autom Lett* 6(1):111–118
- Garcia M, Chatterjee A, Ruina A, Coleman M (1998) The simplest walking model: stability, complexity, and scaling. *J Biomech Eng* 120(2):281–288
- Garcia M, Chatterjee A, Ruina A (2000) Efficiency, speed, and scaling of two-dimensional passive-dynamic walking. *Dyn Stab Syst* 15(2):75–99
- Gardini L, Avrutin V, Sushko I (2014) Codimension-2 border collision, bifurcations in one-dimensional, discontinuous piecewise smooth maps. *Int J Bifurcat Chaos Appl Sci Eng* 24(2):1450024, 30 pages
- Gidoni P (2018) Rate-independent soft crawlers. *Q J Mech Appl Math* 71(4):369–409
- Gidoni P, DeSimone A (2017) Stasis domains and slip surfaces in the locomotion of a bio-inspired two-segment crawler. *Meccanica* 52(3):587–601
- P. Gidoni, F. Riva (2021) A vanishing inertia analysis for finite dimensional rate-independent systems with non-autonomous dissipation, and an application to soft crawlers. arXiv preprint. <https://arxiv.org/abs/2007.09069>
- Glendinning P (2015) Bifurcation from stable fixed point to N-dimensional attractor in the border collision normal form. *Nonlinearity* 28(10):3457–3464
- Glendinning P (2016) Bifurcation from stable fixed point to 2D attractor in the border collision normal form. *IMA J Appl Math* 81(4):699–710
- Glendinning P, Jeffrey MR (2015) Grazing-sliding bifurcations, border collision maps and the curse of dimensionality for piecewise smooth bifurcation theory. *Nonlinearity* 28(1):263–283
- Glendinning P, Kowalczyk P, Nordmark AB (2012) Attractors near grazing-sliding bifurcations. *Nonlinearity* 25(6):1867–1885
- Glendinning P, Kowalczyk P, Nordmark AB (2016) Multiple attractors in grazing-sliding bifurcations in Filippov-type flows. *IMA J Appl Math* 81(4):711–722
- Glover J, Lazer AC, McKenna PJ (1989) Existence and stability of large scale nonlinear oscillations in suspension bridges. *Z Angew Math Phys* 40(2): 172–200
- Goebel R, Sanfelice RG, Teel AR (2009) Hybrid dynamical systems: robust stability and control for systems that combine continuous-time and discrete-time dynamics. *IEEE Control Syst Mag* 29(2):28–93
- Goebel R, Sanfelice RG, Teel AR (2012) Hybrid dynamical systems. Modeling, stability, and robustness. Princeton University Press, Princeton, p xii+212
- Goswami A, Thuijlt B, Espiau B (1998) A study of the passive gait of a compass-like biped robot: symmetry and chaos. *Int J Robot Res* 17(12):1282–1301
- Graef JR, Henderson J, Ouahab A (2013) Impulsive differential inclusions. A fixed point approach, De Gruyter series in nonlinear analysis and applications, vol 20. De Gruyter, Berlin, p xii+400
- Gregg RD, Tilton AK, Candido S, Bretl T, Spong MW (2012) Control and planning of 3-D dynamic walking with asymptotically stable gait primitives. *IEEE Trans Robot* 28(6):1415–1423

- Gritli H, Belghith S (2016) Identification, stability and stabilization of limit cycles in a compass-gait biped model via a hybrid Poincaré map. In: Vaidyanathan S, Volos C (eds) *Advances and applications in nonlinear control systems, Studies in computational intelligence*, vol 635. Springer, Cham, pp 259–289
- Grizzle JW, Abba G, Plestan F (2001) Asymptotically stable walking for biped robots: analysis via systems with impulse effects. *IEEE Trans Autom Control* 46(1):51–64
- Grizzle JW, Chevallereau C, Sinnet RW, Ames AD (2014) Models, feedback control, and open problems of 3D bipedal robotic walking. *Automatica* 50(8):1955–1988
- Guardia M, Seara TM, Teixeira MA (2011) Generic bifurcations of low codimension of planar Filippov systems. *J Differ Equ* 250(4):1967–2023
- Guckenheimer J, Holmes P (1990) *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. Applied mathematical sciences. Revised and corrected reprint of the 1983 original*, vol 42. Springer, New York, p xvi+459
- Heemels WPMH, Brogliato B (2003) The complementarity class of hybrid dynamical systems. *Eur J Control* 9(2–3):322–360
- Holm JK, Spong MW (2008) Kinetic energy shaping for gait regulation of underactuated bipeds. In: 17th IEEE international conference on control applications, pp 1232–1238
- Hu Y, Yan G, Lin Z (2011) Gait generation and control for biped robots with underactuation degree one. *Automatica* 47(8):1605–1616
- Iannelli L, Johansson KH, Jonsson UT, Vasca F (2006) Averaging of nonsmooth systems using dither. *Automatica* 42(4):669–676
- Iida F, Tedrake R (2010) Minimalistic control of biped walking in rough terrain. *Auton Robot* 28(3):355–368
- Ikedo F, Toyama S (2015) A proposal of right and left turning mechanism for quasi-passive walking robot. In: 2015 International conference on advanced robotics and intelligent systems (ARIS), pp 1–5
- Ivanov AP (1994) Impact oscillations: linear theory of stability and bifurcations. *J Sound Vib* 178(3):361–378
- Ivanov AP (1996) Bifurcations in impact systems. *Chaos Solit Fractals* 7(10):1615–1634
- Jones CKRT (1995) Geometric singular perturbation theory. In: *Dynamical systems (Montecatini Terme, 1994)*, Lecture notes in mathematics, vol 1609. Springer, Berlin, pp 44–118
- Kamenskii M, Makarenkov O, Nistri P (2011) Paolo An alternative approach to study bifurcation from a limit cycle in periodically perturbed autonomous systems. *J Dyn Diff Equat* 23(3):425–435
- Kerimoglu D, Karkoub M, Ismail U, Morgul O, Saranlı U (2021) Efficient bipedal locomotion on rough terrain via compliant ankle actuation with energy regulation. *Bioinspir Biomim* 16:056011
- Kolathaya S, Ames AD (2017) Parameter to state stability of control Lyapunov functions for hybrid system models of robots. *Nonlinear Anal Hybrid Syst* 25:174–191
- Kowalczyk P, di Bernardo M, Champneys AR, Hogan SJ, Homer M, Piiroinen PT, Kuznetsov YA, Nordmark A (2006) Two-parameter discontinuity-induced bifurcations of limit cycles: classification and open problems. *Int J Bifurcat Chaos Appl Sci Engrg* 16(3):601–629
- Krasnoselskii MA (1968) The operator of translation along the trajectories of differential equations. *Translations of mathematical monographs*, vol 19. Translated from the Russian by Scripta Technica. American Mathematical Society, Providence, p vi+294
- Kryzhevich SG, Wiercigroch M (2012) Topology of vibro-impact systems in the neighborhood of grazing. *Phys D* 241(22):1919–1931
- Kumar S, Makarenkov O, Gregg RD, Gans N (2022) Stability of time-invariant extremum seeking control for limit cycle minimization, *IEEE Transactions on Automatic Control*, Early Access. <https://doi.org/10.1109/tac.2022.3181253>
- Kupper T, Moritz S (2001) Generalized Hopf bifurcation for non-smooth planar systems. *Non-smooth mechanics. Philos Trans R Soc Lond Ser A Math Phys Eng Sci* 359(1789):2483–2496
- Leine RI, Nijmeijer H (2004) Dynamics and bifurcations of non-smooth mechanical systems, *Lecture notes in applied and computational mechanics*, vol 18. Springer, Berlin, p xii+236
- Levinson N (1950) Small periodic perturbations of an autonomous system with a stable orbit. *Ann Math* 52(2):727–738
- Li T, Chen X (2020) Degenerate grazing-sliding bifurcations in planar Filippov systems. *J Differ Equ* 269(12):11396–11434
- Liberzon D (2003) *Switching in systems and control. In: Systems & control: foundations & applications*. Birkhauser Boston, Inc, Boston, p xiv+233
- Liu L, Tian Y, Huang X (2008) A method to estimate the basin of attraction of the system with impulse effects: application to the biped robots. In: *International conference on intelligent robotics and applications*. Springer, Berlin, pp 953–962
- Liu CJ, Yang J, An K, Liu M, Chen QJ (2020) Robust control of semi-passive biped dynamic locomotion based on a discrete control Lyapunov function. *Robotica* 38(8):1345–1358
- Loud WS (1959) Periodic solutions of a perturbed autonomous system. *Ann Math* 70(2):490–529
- Makarenkov O (2017a) Bifurcation of limit cycles from a fold-fold singularity in planar switched systems. *SIAM J Appl Dyn Syst* 16(3):1340–1371
- Makarenkov O (2017b) A new test for stick-slip limit cycles in dry-friction oscillators with a small nonlinearity in the friction characteristic. *Meccanica* 52(11–12):2631–2640
- Makarenkov O (2019) Bifurcation of limit cycles from a switched equilibrium in planar switched systems. *J Frankl Inst* 356(12):6419–6432

- Makarenkov O (2020) Existence and stability of limit cycles in the model of a planar passive biped walking down a slope. *Proc A* 476(2233):20190450, 17 pages
- Makarenkov O, Lamb JSW (2012) Dynamics and bifurcations of nonsmooth systems: a survey. *Phys D* 241(22): 1826–1844
- Makarenkov O, Niwanthi Wadippuli Achchige L (2018) Bifurcations of finite-time stable limit cycles from focus boundary equilibria in impacting systems, Filippov systems, and sweeping processes. *Int J Bifurcat Chaos Appl Sci Engrg* 28(10):1850126, 13 pages
- Makarenkov O, Ortega R (2011) Asymptotic stability of forced oscillations emanating from a limit cycle. *J Differ Equ* 250(1):39–52
- Makarenkov O, Verhulst F (2021) Resonant periodic solutions in regularized impact oscillator. *J Math Anal Appl* 499(2):125035, 17 pages
- Malkin IG (1956) Some problems in the theory of nonlinear oscillations. Translation of “Nekotorye zadachi teorii nelineinykh kolebaniy” (1956) by Atomic Energy commission, AEC-tr – 3766 books I and II (1959)
- Manchester IR, Mettin U, Iida F, Tedrake R (2011) Stable dynamic walking over uneven terrain. *Int J Robot Res* 30(3):265–279
- McGeer T (1990) Passive dynamic walking. *Int J Robot Res* 9:62–82
- Melnikov VK (1963) On the stability of a center for time-periodic perturbations. *Trudy Moskov Mat Obshch* 12: 3–52. (Russian)
- Misra S, Dankowicz H, Paul MR (2010) Degenerate discontinuity-induced bifurcations in tapping-mode atomic-force microscopy. *Phys D* 239(1–2):33–43
- Newman J, Makarenkov O (2015) Resonance oscillations in a mass-spring impact oscillator. *Nonlinear Dyn* 79(1):111–118
- Nordmark AB (1991) Non-periodic motion caused by grazing incidence in impact oscillators. *J Sound Vib* 2:279–297
- Nordmark AB, Kowalczyk PA (2006) A codimension-two scenario of sliding solutions in grazing-sliding bifurcations. *Nonlinearity* 19(1):1–26
- Obayashi I, Aoi S, Tsuchiya K, Kokubu H (2016) Formation mechanism of a basin of attraction for passive dynamic walking induced by intrinsic hyperbolicity. *Proc R Soc A* 472(2190):20160028
- Or Y (2014) Painlevé’s paradox and dynamic jamming in simple models of passive dynamic walking. *Regul Chaotic Dyn* 19(1):64–80
- Pankov S (2021) Three-dimensional bipedal model with zero-energy-cost walking. *Phys Rev E* 103(4):043003, 18 pages
- Perestyuk NA, Plotnikov VA, Samoilenko AM, Skripnik NV (2011) Differential equations with impulse effects. Multivalued right-hand sides with discontinuities, De Gruyter studies in mathematics, vol 40. Walter de Gruyter & Co., Berlin, p xiv+307
- Perram JW, Shiriaev A, Canudas de Wit C, Grogard F (2003) Explicit formula for a general integral of motion for a class of mechanical systems subject to holonomic constraint. (English summary) *Lagrangian and Hamiltonian methods for nonlinear control* 2003, 8792. IFAC, Laxenburg
- Piiroinen PT, Dankowicz HJ, Nordmark AB (2003) Breaking symmetries and constraints: transitions from 2D to 3D in passive walkers. *Multibody Syst Dyn* 10(2): 147–176
- Reher JP, Hereid A, Kolathaya S, Hubicki CM, Ames AD (2020) Algorithmic foundations of realizing multi-contact locomotion on the humanoid robot DURUS. In: Goldberg K, Abbeel P, Bekris K, Miller L (eds) *Algorithmic foundations of robotics XII*, Springer proceedings in advanced robotics, vol 13. Springer, Cham
- Rom-Kedar V, Turaev D (1999) Big islands in dispersing billiard-like potentials. *Phys D* 130(3–4):187–210
- Sabaapour MR, Yazdi MRH, Beigzadeh B (2015) Passive turning motion of 3D rimless wheel: novel periodic gaits for bipedal curved walking. *Adv Robot* 29(5): 375–384
- Safa AT, Mohammadi S, Hajmiri SE, Naraghi M, Alasty A (2016) How local slopes stabilize passive bipedal locomotion? *Mech Mach Theory* 100:63–82
- Saglam CO, Teel AR, Byl K (2014) Lyapunov-based versus Poincaré map analysis of the rimless wheel. In: 2014 IEEE 53rd annual conference on decision and control (CDC), IEEE, pp 1514–1520
- Sanders JA, Verhulst F, Murdock J (2007) Averaging methods in nonlinear dynamical systems, Applied mathematical sciences, vol 59, 2nd edn. Springer, New York, p xxii+431
- Santisteban R, Fridman L, Moreno J (2010) Finite-time convergence analysis for twisting controller via a strict Lyapunov function. In: Proceedings of 2010 11th international workshop on variable structure systems, Mexico City, 26–28 June 2010
- Schwab A, Wisse M (2001) Basin of attraction of the simplest walking model. In: Proceedings of ASME 2001 design engineering technical conferences and computers and information in engineering conference, pp 1–9
- Sfacci A (2017) Periodic impact motions at resonance of a particle bouncing on spheres and cylinders. *Adv Nonlinear Stud* 17(3):481–496
- Shiriaev AS, Freidovich LB, Gusev SV (2010) Transverse linearization for controlled mechanical systems with several passive degrees of freedom. *IEEE Trans Autom Control* 55(4):893–906
- Sidorov E, Zacksenhouse M (2019) Lyapunov based estimation of the basin of attraction of Poincaré maps with applications to limit cycle walking. *Nonlinear Anal Hybrid Syst* 33:179–194
- Simic SN, Johansson KH, Lygeros J, Sastry S (2005) Towards a geometric theory of hybrid systems. *Dyn Contin Discrete Impuls Syst Ser B Appl Algorithm* 12(5–6):649–687
- Simpson DJW (2010) Bifurcations in piecewise-smooth continuous systems, World scientific series on nonlinear science. Series A: monographs and treatises,

- vol 70. World Scientific Publishing Co. Pte. Ltd, Hackensack, p xv+238
- Simpson DJW (2017) Grazing-sliding bifurcations creating infinitely many attractors. *Int J Bifurcat Chaos Appl Sci Engrg* 27(12):1730042, 15 pages
- Simpson DJW (2018) A compendium of Hopf-like bifurcations in piecewise-smooth dynamical systems. *Phys Lett A* 382(35):2439–2444
- Simpson DJW, Meiss JD (2007) Andronov-Hopf bifurcations in planar, piecewise-smooth, continuous flows. *Phys Lett A* 371(3):213–220
- Simpson DJW, Meiss JD (2012) Aspects of bifurcation theory for piecewise-smooth, continuous systems. *Phys D* 241(22):1861–1868
- Smith AC, Berkemeier MD (1998) The motion of a finite-width rimless wheel in 3D. In: *Proceedings of IEEE international conference on robotics and automation*, vol 1–4, pp 2345–2350
- Sobajima M, Kobayashi T, Sekiyama K, Fukuda T (2013) Bipedal walking control of humanoid robots by arm-swing. In: *Proceedings of the SICE annual conference*, pp 313–318
- Sovero SE, Saglam CO, Byl K (2015) Passive frontal plane coupling in 3D walking. In: *IEEE/RSJ international conference on intelligent robots and systems*, pp 1605–1611
- Spong MW (1999) Passivity based control of the compass gait biped. *IFAC Proc* 32(2):506–510
- Spong M, Bullo F (2005) Controlled symmetries and passive walking. *IEEE Trans Autom Control* 50(7):1025–1031
- Spong MW, Holm JK, Lee D (2007) Passivity-based control of bipedal locomotion – regulating walking by exploiting passive gaits in 2-D and 3-D bipeds. *IEEE Robot Autom Mag* 14(2):30–40
- Spong MW, Hutchinson S, Vidyasagar M (2020) *Robot modeling and control*, 2nd edn. Wiley, Hoboken, p 608
- Su JL, Dingwell JB (2007) Dynamic stability of passive dynamic walking on an irregular surface. *Trans ASME* 129:802–810
- Sushko I, Gardini L (2010) Degenerate bifurcations and border collisions in piecewise smooth 1D and 2D maps. *Int J Bifurcat Chaos Appl Sci Engrg* 20(7):2045–2070
- Szalai R, Osinga HM (2008a) Unstable manifolds of a limit cycle near grazing. *Nonlinearity* 21(2):273–284
- Szalai R, Osinga HM (2008b) Invariant polygons in systems with grazing-sliding. *Chaos* 18(2):023121, 11 pages
- Szalai R, Osinga HM (2009) Arnold tongues arising from a grazing-sliding bifurcation. *SIAM J Appl Dyn Syst* 8(4):1434–1461
- Tang JZ, Boudali AM, Manchester IR (2018) Invariant funnels for underactuated dynamic walking robots: new phase variable and experimental validation. In: *Proc. IEEE international conference on robotics and automation*, pp 3497–3504
- Tedrake R, Zhang TW, Fong MF, Seung HS (2004) Actuating a simple 3D passive dynamic walker. In: *Proceedings of 2004 IEEE international conference on robotics and automation (ICRA)*, vol 1–5, pp 4656–4661
- Thomsen JJ, Fidler A (2008) Near-elastic vibro-impact analysis by discontinuous transformations and averaging. *J Sound Vib* 311(1–2):386–407
- Turaev D, Rom-Kedar V (1998) Elliptic islands appearing in near-ergodic flows. *Nonlinearity* 11(3):575–600
- Veer S, Poulakakis I (2020) Switched systems with multiple equilibria under disturbances: boundedness and practical stability. *IEEE Trans Autom Control* 65(6):2371–2386
- Veer S, Motahar MS, Poulakakis I (2017) Almost driftless navigation of 3D limit-cycle walking bipeds. In: *IEEE/RSJ international conference on intelligent robots and systems*, pp 5025–5030
- Verhulst F (2005) *Methods and applications of singular perturbations, boundary layers and multiple timescale dynamics*. Springer, New York
- Wang Q, Wang L, Huang Y, Zhu J, Chen W (2009) Three-dimensional quasi-passive dynamic bipedal walking with flat feet and compliant ankles. In: *Proceedings of the 48th IEEE conference on decision and control jointly with the 28th Chinese control conference (CDC/CCC 2009)*, pp 8200–8205
- Weiss D, Kupper T, Hosham HA (2012) Invariant manifolds for nonsmooth systems. *Phys D* 241(22):1895–1902
- Westervelt ER, Grizzle JW, Koditschek DE (2003) Hybrid zero dynamics of planar biped walkers. *IEEE Trans Autom Control* 48(1):42–56
- Westervelt ER, Buche G, Grizzle JW (2004) Experimental validation of a framework for the design of controllers that induce stable walking in planar bipeds. *Int J Robot Res* 23(6):559–582
- Wisse M, Schwab AL, vd. Linde RQ (2001) A 3D passive dynamic biped with yaw and roll compensation. *Robotica* 19(3):275–284
- Wisse M, Schwab AL, van der Linde RQ, van der Helm FCT (2005) How to keep from falling forward: elementary swing leg action for passive dynamic walkers. *IEEE Trans Robot* 21(3):393–401
- Yagasaki K (2004) Nonlinear dynamics of vibrating micro-cantilevers in tapping mode atomic force microscopy. *Phys Rev B* 70:245–419
- Yang T, Westervelt ER, Serrani A, Schmiedeler JP (2009) A framework for the control of stable aperiodic walking in underactuated planar bipeds. *Auton Robot* 27(3):277–290
- Ye H, Michel AN, Hou L (1998) Stability theory for hybrid dynamical systems. *Hybrid control systems. IEEE Trans Autom Control* 43(4):461–474
- Young LS (1979) A closing lemma on the interval. *Invent Math* 54(2):179–187
- Yudaev SA, Rachinskii D, Vladimir A (2017) An asymptotic solution to a passive biped walker model. *J Phys Conf Ser* 811:012018, 7 pages
- Zhai G, Hu B, Yasuda K, Michel AN (2001) Stability analysis of switched systems with stable and unstable subsystems: an average dwell time approach. *Int J Syst Sci* 32(8):1055–1061
- Zhang Y, Makarenkov O, Gans N (2016) Extremum seeking control of a nonholonomic system with sensor constraints. *Automatica* 70:86–93

- Zhao X, Dankowicz H (2006) Unfolding degenerate grazing dynamics in impact actuators. *Nonlinearity* 19(2): 399–418
- Zhao D, Revzen S (2020) Multi-legged steering and slipping with low DoF hexapod robots. *Bioinspir Biomim* 15:045001
- Znegui W, Gritli H, Belghith S (2020a) Design of an explicit expression of the Poincaré map for the passive dynamic walking of the compass-gait biped model. *Chaos Solit Fractals* 130:109436, 18 pages
- Znegui W, Gritli H, Belghith S (2020b) Stabilization of the passive walking dynamics of the compass-gait biped robot by developing the analytical expression of the controlled Poincaré map. *Nonlinear Dyn* 101(2):1061–1091
- Zou Y, Kupper T, Beyn W-J (2006) Generalized Hopf bifurcation for planar Filippov systems continuous at the origin. *J Nonlinear Sci* 16(2):159–177

# Perturbation Theory for Water Waves

Riccardo Montalto  
Dipartimento di Matematica “Federigo Enriques”,  
Università degli Studi di Milano, Milan, Italy

## Article Outline

Glossary  
Definition of the Subject  
Introduction  
KAM Results for Water Waves  
Longtime Existence for Periodic Water Waves  
Future Developments  
Bibliography

## Glossary

**KAM theory and quasi-periodic solutions** The theory, introduced by Kolmogorov, Arnold and Moser, which allows to study the existence and the stability properties of *quasi-periodic solutions* for a given nearly integrable dynamical system. More precisely, let us consider a dynamical system (even infinite dimensional)  $\partial_t u = X(u)$ ,  $u \in \mathcal{H}$  where  $\mathcal{H}$  is a suitable Hilbert space. A quasi-periodic solution of this dynamical system is a solution of the form  $u(t) := U(\omega t)$  where  $U(\varphi_1, \dots, \varphi_\nu)$  is a smooth function, periodic in all its variables and  $\omega := (\omega_1, \dots, \omega_\nu) \in \mathbb{R}^\nu$  is a nonresonant frequency vector, that is,  $\omega \cdot \ell \neq 0$ ,  $\forall \ell \in \mathbb{Z}^{\nu} \setminus \{0\}$ .

**Normal form** A procedure, possibly iterative, which allows transforming an equation or an operator in a form which is easier to be analyzed. For instance, in the framework of linear operators which are small perturbations of

diagonal operators, a possible normal form procedure is the construction of a change of coordinates, which reduces the operator under consideration to a diagonal one. This is usually possible only under certain technical conditions. (cf. also the article Gaeta (2014)).

**Perturbation theory** In the framework of non-linear differential equations, Perturbation theory is a collection of methods allowing to study solutions nearby a known solutions (or approximate solutions), like equilibria. In the framework of linear operators which are (for instance) an  $\varepsilon$ -perturbation of a diagonal one, it is the theory which allows determining the spectrum as a power series in the small parameter  $\varepsilon$  (cf. also the article Gaeta (2014)).

**Water Waves equations** The equations which describe the motion of a perfect, incompressible, irrotational fluid, under the action of the gravity, which occupies the free boundary region  $\mathcal{D}_\eta := \{(x, y) \in \Omega \times \mathbb{R} : -h < y < \eta(t, x)\}$  where  $h > 0$ ,  $\Omega$  is a subdomain of  $\mathbb{R}^d$  and  $\eta : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$  is a smooth function to be determined. The unknowns of the problem are the function  $\eta$  and the velocity field  $(t, x, y) \in \mathbb{R} \times \mathcal{D}_\eta \mapsto u(t, x, y) \in \mathbb{R}^d$  which is the gradient of a scalar function  $\Phi(t, x, y)$ , called velocity potential (see section “Definition of the Subject” below).

## Definition of the Subject

The celebrated Water Waves equation describes the motion of a perfect, incompressible, irrotational fluid, which occupies the free boundary region

$$\mathcal{D}_\eta := \{(x, y) \in \Omega \times \mathbb{R} : -h < y < \eta(t, x)\}$$

under the action of the gravity, where  $\Omega$  is a subdomain of  $\mathbb{R}^d$ . In this article, we focus on the cases where



$$\Omega = \mathbb{T}^d \quad (1)$$

where  $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$  is a one dimensional torus. Clearly, the physical situations are  $d = 2$  (three dimensional Water Waves) and  $d = 1$  (two-dimensional Water Waves). Since the fluid is incompressible and irrotational, the velocity field in the region  $\mathcal{D}_\eta$  is the gradient of a function  $\Phi(t, x, y)$ , which we call *velocity potential*. The time evolution of the fluid is determined by the two functions  $(t, x) \mapsto \eta(t, x)$ ,  $(t, x, y) \mapsto \Phi(t, x, y)$ . Since  $\eta$  is an unknown of the problem, then the domain  $\mathcal{D}_\eta$  is also unknown. This is the reason why the Water Waves problem is a free boundary problem. More precisely, the dynamics of the system is described by the equations

$$\begin{cases} \partial_t \Phi + \frac{1}{2} |\nabla \Phi|^2 + g\eta = \kappa \operatorname{div} \left( \frac{\nabla \eta}{\sqrt{1 + |\nabla \eta|^2}} \right) \\ \quad \text{at } y = \eta(x) \\ \Delta \Phi = 0 \text{ in } \mathcal{D}_\eta \\ \partial_y \Phi = 0 \text{ at } y = -h \\ \partial_t \eta = \partial_y \Phi - \nabla \eta \cdot \nabla_x \Phi \text{ at } y = \eta(x) \end{cases} \quad (2)$$

where  $g$  is the acceleration of gravity,  $\kappa \in [\kappa_1, \kappa_2]$ ,  $\kappa_1 > 0$ , is the surface tension coefficient and

$$\operatorname{div} \left( \frac{\nabla \eta}{\sqrt{1 + |\nabla \eta|^2}} \right)$$

is the mean curvature of the free surface. The unknowns of the problem are the free surface  $y = \eta(x)$  and the velocity potential  $\Phi : \mathcal{D}_\eta \rightarrow \mathbb{R}$ , that is, the irrotational velocity field  $v = \nabla_{x,y} \Phi$  of the fluid. The first equation in eq. (2) is the Bernoulli condition according to which the jump of pressure across the free surface is proportional to the mean curvature. The last equation (in eq. (2)) expresses that the velocity of the free surface

coincides with one of the fluid particles. In the case where the depth of the ocean  $\mathcal{D}_\eta$  is infinite, that is,  $h = \infty$ , then the third equation (in eq. (2)) is replaced by the condition

$$\partial_y \Phi \rightarrow 0 \text{ as } y \rightarrow -\infty. \quad (3)$$

For the physical derivation of the water waves problem and their approximate models, we refer to the classical monographs Whitham (1999), Stoker (1992) (see also the more recent monograph Lannes (2013)).

## Introduction

The periodic Water Waves equations have been extensively studied from the point of view of Hamiltonian perturbation theory. It was actually proved by Zakharov (1968) and Craig and Sulem (1993) that the evolution problem eq. (2) may be written as an infinite dimensional Hamiltonian system (we refer to the monograph Lannes (2013) for the Lagrangian formulation of the Water Waves problem). At each time  $t \in \mathbb{R}$  the profile  $\eta(t, x)$  of the fluid and the value

$$\psi(t, x) = \Phi(t, x, \eta(t, x))$$

of the velocity potential  $\Phi$  restricted to the free boundary uniquely determine the velocity potential  $\Phi$  in the whole  $\mathcal{D}_\eta$ , solving (at each  $t$ ) the elliptic problem (see e.g., Alazard et al. (2011), Lannes (2013))

$$\begin{cases} \Delta \Phi = 0 \text{ in } \mathcal{D}_\eta, \Phi|_{y=\eta} = \psi, \\ \partial_y \Phi(x, y) = 0 \text{ at } y = -h \end{cases} \quad (4)$$

and in the case of infinite depth

$$\begin{cases} \Delta \Phi = 0 \text{ in } \mathcal{D}_\eta, \Phi|_{y=\eta} = \psi, \\ \partial_y \Phi(x, y) \rightarrow 0 \text{ as } y \rightarrow -\infty. \end{cases} \quad (5)$$

As proved in Zakharov (1968), Craig and Sulem (1993), system eq. (2) is then equivalent to the system

$$\begin{cases} \partial_t \eta = G(\eta) \psi, \\ \partial_t \psi + \eta + \frac{1}{2} |\nabla \psi|^2 - \frac{1}{2} \frac{(G(\eta) \psi + \nabla \psi \cdot \nabla \eta)}{1 + |\nabla \eta|^2} = \\ \kappa \operatorname{div} \frac{\nabla \eta}{(1 + |\nabla \eta|^2)^{\frac{1}{2}}} \end{cases} \quad (6)$$

where  $G(\eta)$  is the so-called Dirichlet-Neumann operator defined by

$$\begin{aligned} G(\eta) \psi(x) &:= \sqrt{1 + \eta_x^2} \partial_n \Phi|_{y=\eta(x)} \\ &= (\partial_y \Phi)(x, \eta(x)) - \nabla \eta(x) \cdot (\nabla_x \Phi) \\ &\quad \times (x, \eta(x)) \end{aligned} \quad (7)$$

The operator  $G(\eta)$  is linear in  $\psi$ , self-adjoint with respect to the  $L^2$  (with respect to the  $x$ -variable) scalar product and semi positive definite. The Kernel of  $G(\eta)$  is the one-dimensional space of the constant functions.

Furthermore, the eq. (21) are the Hamiltonian system (see Zakharov (1968), Craig and Sulem (1993))

$$\begin{aligned} \partial_t \eta &= \nabla_\psi H(\eta, \psi), \partial_t \psi = -\nabla_\eta H(\eta, \psi) \\ \partial_t u &= J \nabla_u H(u), u = \begin{pmatrix} \eta \\ \psi \end{pmatrix}, J := \begin{pmatrix} 0 & \operatorname{Id} \\ -\operatorname{Id} & 0 \end{pmatrix}, \end{aligned} \quad (8)$$

where  $\nabla$  denotes the  $L^2$ -gradient, and the Hamiltonian

$$\begin{aligned} H(\eta, \psi) &:= \frac{1}{2} (\psi, G(\eta) \psi)_{L^2} + g \int_{\mathbb{T}^d} \frac{\eta^2}{2} dx \\ &\quad + \kappa \int_{\mathbb{T}^d} \sqrt{1 + |\nabla \eta|^2} dx \end{aligned} \quad (9)$$

is the sum of the kinetic energy

$$K := \frac{1}{2} (\psi, G(\eta) \psi)_{L^2} = \frac{1}{2} \int_{\mathcal{D}_\eta} |\nabla \Phi|^2(x, y) dx dy,$$

the potential energy

$$W := \int_{\mathbb{T}^d} \eta^2 dx$$

the energy of the capillary forces (area surface integral)

$$T := \kappa \int_{\mathbb{T}^d} \sqrt{1 + |\nabla \eta|^2} dx$$

expressed in terms of the variables  $(\eta, \psi)$ .

The symplectic structure induced (by eq. 8) is the standard Darboux 2-form

$$\begin{aligned} \mathcal{W}(u_1, u_2) &:= (u_1, Ju_2)_{L^2} \\ &= (\eta_1, \psi_2)_{L^2} - (\psi_1, \eta_2)_{L^2} \end{aligned} \quad (10)$$

for all  $u_1 = (\eta_1, \psi_1), u_2 = (\eta_2, \psi_2)$ .

We use the following terminology. We refer (to eq. 21) when  $\kappa \neq 0$  as *gravity-capillary Water Waves equations* and we refer (to eq. 21) when  $\kappa = 0$  as *pure gravity Water Waves equations*. Moreover, when  $h = +\infty$ , we refer (to eq. 21) as *Water Waves equations infinite depth*.

In this paper we shall present some *perturbative results* on the Water Waves equation, namely we shall analyze the behavior of the system eq. (6), close to the origin  $(\eta, \psi) = (0, 0)$ . This means essentially to look for solutions having small velocity and small wave amplitude. Hence, the first thing to look at is the linearization of the equation at the origin.

It turns out that the linearized equation at  $(\eta, \psi) = (0, 0)$  takes the form

$$\begin{cases} \partial_t \eta = |D| \tanh(hD) [\psi] \\ \partial_t \psi = -(g - \kappa \Delta) \eta \end{cases} \quad (11)$$

where the *Fourier multiplier*  $|D| \tanh(hD)$  is the Fourier diagonal operator given by

$$\begin{aligned} |D| \tanh(hD) [\psi] &:= \sum_{\xi \in \mathbb{Z}^d} |\xi| \tanh(h\xi) \widehat{\psi}(\xi) e^{i x \cdot \xi} \text{ where} \\ \psi(x) &= \sum_{\xi \in \mathbb{Z}^d} \widehat{\psi}(\xi) e^{i x \cdot \xi} \in L^2(\mathbb{T}^d). \end{aligned}$$

(12)

By introducing the complex valued unknown

$$\begin{aligned} u &:= \left( \sqrt{g - \Delta} \right) [\eta] + i \sqrt{|D| \tanh(hD)} [\psi] \\ &= \sum_{\xi \in \mathbb{Z}^d} \sqrt{g + |\xi|^2} \widehat{\eta}(\xi) e^{i x \cdot \xi} \\ &\quad + i \sum_{\xi \in \mathbb{Z}^d} \sqrt{|\xi| \tanh(h\xi)} \widehat{\psi}(\xi) e^{i x \cdot \xi} \end{aligned} \quad (13)$$

the linear eq. (11) is then equivalent to

$$\partial_t u + i \Lambda_{h,\kappa}(D) u = 0 \quad (14)$$

where

$$\Lambda_{h,\kappa}(D) = \sqrt{|D| \tanh(hD) (g - \kappa \Delta)}. \quad (15)$$

The solutions of the eq. (14) are then given by

$$u(t, x) = \sum_{\xi \in \mathbb{Z}^d} e^{-i \Lambda_{h,\kappa}(\xi) t} \widehat{u}(0, \xi) e^{i x \cdot \xi}, \quad (16)$$

which implies that the solutions are superposition's of infinitely many harmonic oscillators, oscillating with frequencies  $\Lambda_{h,\kappa}(\xi) := \sqrt{|\xi| \tanh(h\xi) (g + \kappa |\xi|^2)}$ ,  $\xi \in \mathbb{Z}^2$ . Hence all the solutions are either periodic (oscillating with one frequency) or quasi-periodic (oscillating with finitely many frequencies) or almost periodic (oscillating with infinitely many frequencies). As a consequence all the  $H^s$ -norms are constant for all times where, for any  $s \geq 0$ , we define the Sobolev space  $H^s(\mathbb{T}^d)$  as the space of periodic functions  $f(x)$ , equipped with the norm

$$\|f\|_s := \left( \sum_{\xi \in \mathbb{Z}^d} \langle \xi \rangle^{2s} |\widehat{f}(\xi)|^2 \right)^{\frac{1}{2}}, \quad \langle \xi \rangle := (1 + |\xi|^2)^{\frac{1}{2}}$$

where

$$\widehat{f}(\xi) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x) e^{-i x \cdot \xi} dx, \quad \forall \xi \in \mathbb{Z}^d.$$

In the framework of the Water Waves problem, the purpose of perturbation theory is to investigate

the behavior of the nonlinear equation, regarded as a perturbation of the linearized equation at the origin. The pioneering paper of Stokes (1847) is the starting point for the nonlinear theory of dispersive waves. It was in this work, and far ahead of other developments that he found the crucial results that, first, periodic wavetrains are possible in nonlinear systems and, second that the dispersion relation involves the amplitude. More precisely, Stokes constructed (at least formally) traveling wave solutions for the 2d Water Waves system eq. (2) (the horizontal variable  $x$  is one dimensional), as a power series with respect to the wave amplitude  $\varepsilon$ . He looked for traveling wave solutions of the form

$$(\Phi(kx - \omega t, y), \eta(kx - \omega t)) \quad (17)$$

where the functions  $\Phi(\theta, y)$  and  $\eta(\theta)$  are periodic in  $\theta$  and they admit the expansion

$$\begin{aligned} \Phi(\theta, y) &= \varepsilon \Phi_1(\theta, y) + \varepsilon^2 \Phi_2(\theta, y) + \varepsilon^3 \Phi_3(\theta, y) + \dots \\ \eta(\theta) &= \varepsilon \eta_1(\theta) + \varepsilon^2 \eta_2(\theta) + \varepsilon^3 \eta_3(\theta) + \dots \end{aligned} \quad (18)$$

In order to avoid secular terms, also the frequency  $\omega$  has to be modulated with the amplitude  $\varepsilon$ , namely

$$\omega(k) = \omega_0(k) + \varepsilon \omega_1(k) + \varepsilon^2 \omega_2(k) + \dots \quad (19)$$

(it turns out by calculations that the term  $\varepsilon \omega_1(k)$  vanishes). By the ansatz eqs. (17), (18), (19), one can solve the eq. (2) in any order and determine inductively  $\Phi_1, \Phi_2, \Phi_3, \dots, \eta_1, \eta_2, \eta_3, \dots, \omega_0(k), \omega_2(k), \dots$ . For more details on the formal construction of the Stokes' waves, we refer to the classical monograph Whitham (1999), Chap. 13. The first rigorous construction of small amplitude space periodic steady traveling waves goes back to the 1920s with the papers of Nekrasov (1921), Levi-Civita (1925) and Struik (1926), in the case of pure gravity two-dimensional Water-Waves. Later, Zeidler (1971) considered the effect of capillarity. In the presence of vorticity, the first result is due to Gerstner (1802) in 1802, who gave an explicit example of a periodic traveling wave, in infinite depth, and with a particular nonzero

vorticity. One had to wait for the work of Dubreil-Jacotin (1934) in 1934 for the first existence results of small amplitude, periodic traveling waves with general (Hölder continuous, small) vorticity, and, later, the works of Goyon (1958) and Zeidler (1973) in the case of large vorticity (see also the more recent papers Wahlén (2006), Martin (2013)). All these results deal with 2d water waves, and can ultimately be deduced by the Crandall-Rabinowitz bifurcation theorem from a simple eigenvalue. We also mention that these local bifurcation results can be extended to global branches of steady traveling waves by applying the methods of global bifurcation theory. We refer to Keady and Norbury (1978), Toland (1978), McLeod (1997) and Constantin and Strauss (2004). For a more complete list of references, concerning periodic traveling waves, we refer to the monograph Constantin (2011).

The persistence of periodic traveling wave solutions has been one of the first problems to be investigated at a nonlinear level. The key feature of the periodic traveling wave solutions is that they are *stationary in a moving frame*. This simplifies a lot, the mathematical analysis of the problem, since one essentially solves a steady problem. More in general, when analyzing the dynamics of the nonlinear Water Waves system, the following natural questions arise.

1. **Local well posedness.** The first question that one has to address when dealing with nonlinear Partial differential equations is the local well posedness: do smooth solutions of the nonlinear system exist at least for a short interval of times (and possibly for small data)?
2. **Persistence of oscillatory solutions.** Do periodic, quasi-periodic or almost periodic solutions persist for the nonlinear system at least for suitable choices of initial data and parameters ( $\kappa$  and  $\mathbf{h}$ )? The study of this particular solution in the framework of infinite dimensional dynamical systems is referred to as the KAM (Kolmogorov-Arnold-Moser) theory for PDEs.
3. **Long time dynamics.** For small initial data  $O(\varepsilon)$ , is the corresponding solution small of order  $O(\varepsilon)$  for large times or eventually globally in time?

The main general difficulty in addressing the aforementioned problems for the Water Waves is the **fully nonlinear nature** of the Water Waves equation. This means that the nonlinear part of the equation contains as many derivatives as its linear part. More precisely this means that in the complex unknown introduced in eq. (13), the Water Waves eq. (6) takes the form

$$\partial_t u + i\Lambda_{h,\kappa}(D)u + Q(u) = 0 \quad (20)$$

where  $Q(u)$  is quadratic in  $u$  and it is an unbounded nonlinear operator of order  $3/2$ , which is the same order of the Fourier multiplier  $\Lambda_{h,\kappa}(D)$ . The general technical tool for dealing with fully nonlinear PDE is given by microlocal analysis and in particular by pseudo-differential calculus and its generalization to symbols with limited smoothness, which is called Paradifferential calculus. We refer to Taylor (1991), Saranen and Vainikko (2002) for a complete overview on the theory of pseudo-differential calculus and to Metivier (2008) for paradifferential calculus. For clarity of the reader, we have given the precise definition of periodic pseudo-differential operators, since they are used throughout this paper.

**Definition 3.1** Let  $m \in \mathbb{R}$ . A  $C^\infty$  function  $a : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$  is in the class  $S^m$  if and only if for any  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d, \beta = (\beta_1, \dots, \beta_d) \in \mathbb{N}^d$ , there exists a constant  $C(\alpha, \beta) > 0$  such that

$$\left| \partial_x^\alpha \partial_\xi^\beta a(x, \xi) \right| \leq C(\alpha, \beta) \langle \xi \rangle^{m-|\beta|}, \forall (x, \xi) \in \mathbb{T}^d \times \mathbb{R}^d,$$

where  $\langle \xi \rangle := (1 + |\xi|^2)^{\frac{1}{2}}$ . We then say that a linear operator  $A$  is in the class  $OP S^m$  if there exists a symbol  $a \in S^m$  such that

$$\begin{aligned} Au(x) &= \sum_{\xi \in \mathbb{Z}^d} a(x, \xi) \widehat{u}(\xi) e^{ix \cdot \xi} \text{ where } u(x) \\ &= \sum_{\xi \in \mathbb{Z}^d} \widehat{u}(\xi) e^{ix \cdot \xi}. \end{aligned}$$

We use the notation  $A = Op(a)$ .

Note that the fully nonlinear nature of the equation makes nontrivial even the problem of local well-posedness. On the other hand, the local theory is nowadays well established and it has been addressed by many authors, see for example Nalimov (1974), Yosihara (1982), Craig (1985), Wu (1997), Christodoulou and Lindblad (2000), Shatah and Zeng (2011), Lannes (2005), Alazard et al. (2011), Alazard et al. (2014). We also refer to the review (Ionescu and Pusateri 2018b, Section 2) and to the monograph Lannes (2013) for an extensive list of references. In particular, for smooth enough initial data which are of size  $\varepsilon$ , the solutions exist and stay regular for times of order  $\varepsilon^{-1}$ . We actually give the precise statement of the local well posedness result proved in Alazard-Burq-Zuily (2011) for the gravity capillary Water Waves. The result holds in the case where the horizontal variable  $x$  is in  $\mathbb{T}^d$  or  $\mathbb{R}^d$  for  $d \geq 1$ .

**Theorem 3.2 (Alazard, Burq, Zuily)** *Let  $\Omega$  be  $\mathbb{R}^d$  or  $\mathbb{T}^d$  and  $s > d/2 + 2$ ,  $(\eta_0, \psi_0) \in H^{s+\frac{1}{2}}(\Omega) \times H^s(\Omega)$ . Then, there exists a time  $T > 0$  such that the Cauchy problem eq. (6) with initial data  $(\eta_0, \psi_0)$  admits a unique solution*

$$(\eta, \psi) \in C^0\left([0, T], H^{s+\frac{1}{2}}(\Omega) \times H^s(\Omega)\right).$$

A quite similar statement holds in the case of pure gravity Water Waves, that is,  $\kappa = 0$  (no surface tension), cf. Alazard et al. (2014).

Concerning problems 2 and 3 listed above (in the case of periodic horizontal variable  $x \in \mathbb{T}^d$ ), the situation is much more complicated. We discuss Problem 2 in detail in the next section “KAM Results for Water Waves” and Problem 3 in section “Longtime Existence for Periodic Water Waves.”

## KAM Results for Water Waves

In this section, we shall describe the KAM results which have been obtained for the Water Waves systems. Up to now, these results have been only obtained for 2d fluids, hence, we shall take the horizontal variable  $x \in \mathbb{T}$ , where  $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$  is

the one dimensional torus. The Zakharov, Craig-Sulem formulation, then takes the form

$$\begin{cases} \partial_t \eta = G(\eta) \psi, \\ \partial_t \psi + \eta + \frac{1}{2} \psi_x^2 - \frac{1}{2} \frac{(G(\eta) \psi + \eta_x \psi_x)^2}{1 + \eta_x^2} = \\ \kappa \frac{\eta_{xx}}{(1 + \eta_x^2)^{3/2}}, \end{cases} \quad (21)$$

which is a pseudo-PDE in one dimension (we have taken the gravity  $g = 1$  since it is not relevant). The Dirichlet-Neumann operator  $G(\eta)$  satisfies an important property, which has been proved by Calderon, namely

$$G(\eta) = |D| \tanh(hD) + \mathcal{R}(\eta) \quad (22)$$

where the Fourier multiplier  $|D| \tanh(hD)$  is defined by

$$\begin{aligned} |D| \tanh(hD) u(x) &:= \sum_{\xi \in \mathbb{Z} \setminus \{0\}} \tanh(h\xi) |\xi| \widehat{u}(\xi) e^{i x \xi}, \\ u(x) &= \sum_{\xi \in \mathbb{Z}} \widehat{u}(\xi) e^{i x \xi} \in L^2(\mathbb{T}) \end{aligned}$$

( $\tanh(y)$  is the hyperbolic tangent) and  $\mathcal{R}(\eta)$  is an arbitrarily regularizing remainder, namely

$$\begin{aligned} \mathcal{R}(\eta) : H^s(\mathbb{T}) &\rightarrow H^{s+\sigma}(\mathbb{T}), \quad \forall s, \sigma \\ &\geq 0 \text{ provided } \eta \in C^\infty(\mathbb{T}). \end{aligned}$$

In this Section we shall describe the recent results concerning the existence of periodic and quasi-periodic solutions for eq. (21). In particular, we shall describe in detail the results concerning the existence of quasi-periodic solutions in the gravity capillary case in infinite depth Berti and Montalto (2020) and in the pure gravity case in finite depth Baldi et al. (2018). We recall the definition of quasi-periodic solution. Given a dynamical system (even infinite dimensional)

$$\partial_t u = X(u), \quad u \in \mathcal{H} \quad (23)$$

where  $\mathcal{H}$  is a suitable Hilbert space, a quasi-periodic solution of eq. (23), is a solution of the

form  $u(t) := U(\omega t)$  where  $\omega := (\omega_1, \dots, \omega_v) \in \mathbb{R}^v$  is a non-resonant frequency vector, that is,

$$\omega \cdot \ell \neq 0, \quad \forall \ell \in \mathbb{Z}^v \setminus \{0\}$$

and  $U : \mathbb{T}^v \rightarrow \mathcal{H}$  is a smooth function which is periodic in all its variables. The function  $\varphi \mapsto U(\varphi)$  satisfies then the functional equation

$$\omega \cdot \partial_\varphi U - X(U) = 0. \quad (24)$$

The KAM (Kolmogorov-Arnold-Moser) theory has been introduced in the late 1950s, by Kolmogorov, Arnold, and Moser in the case of finite dimensional Hamiltonian systems and it has been generalized in the late 1980s to partial differential equations by Kuksin (1987), Wayne (1990), Craig-Wayne (1993), Bourgain (1994), and Pöschel (1996). We refer to Berti (2016) for a detailed list of references. The problem of looking for periodic and quasi-periodic solutions of dispersive and hyperbolic PDEs is the *small divisor's* problem. We shall explain below the small divisor's problem in detail. Let us consider a general dispersive one dimensional PDE of the form

$$\partial_t u + i\Lambda(D)u + Q(u) = 0, \quad x \in \mathbb{T} \quad (25)$$

where  $\Lambda(D)$  is a Fourier multiplier of order  $m$  and  $Q$  is a smooth nonlinear operator which is quadratic w.r. to  $u$  and possibly unbounded of order  $\sigma \leq m$ . By rescaling the variables  $u$  it is possible to introduce a small parameter  $\varepsilon$  in front of the nonlinearity and hence, eq. (25) becomes

$$\partial_t u + i\Lambda(D)u + \varepsilon Q(u) = 0, \quad x \in \mathbb{T} \quad (26)$$

The problem of finding quasi-periodic solutions  $u(\omega t, x)$ ,  $\omega \in \mathbb{R}^v$ ,  $u \in H^s(\mathbb{T}^v \times \mathbb{T})$  of the PDE is then equivalent to find zeros of the non-linear operator

$$\begin{aligned} \mathcal{F}(u, \varepsilon) &:= \omega \cdot \partial_\varphi u + i\Lambda(D)u + \varepsilon Q(u) = 0, \\ \mathcal{F} : H^{s+m}(\mathbb{T}^v \times \mathbb{T}) \times [0, \varepsilon_0] &\rightarrow H^s(\mathbb{T}^v \times \mathbb{T}) \end{aligned} \quad (27)$$

One then could try to apply the classical implicit function theorem. Indeed one has that

$\mathcal{F}(0, 0) = 0$  and the Frechet differential (with respect to  $u$ ) at  $(0, 0)$  is given by

$$\begin{aligned} D_u \mathcal{F}(0, 0)h &= (\omega \cdot \partial_\varphi + i\Lambda(D))h \\ &= \sum_{\ell, j} i(\omega \cdot \ell + \Lambda(j)) \widehat{h}(\ell, j) e^{i\ell \cdot \varphi} e^{ijx}. \end{aligned}$$

The point is that the sequence  $\{\omega \cdot \ell + \Lambda(j) : (\ell, j) \in \mathbb{Z}^v \times \mathbb{Z}\}$  (even in the case where  $\omega$  is one-dimensional) accumulates to zero for a.e.  $\omega \in \mathbb{R}^v$ . The best lower bound that can be verified for a *large set* of parameters  $\omega$  is the *diophantine condition*

$$|\omega \cdot \ell + \Lambda(j)| \geq \frac{\gamma}{(\ell)^\tau}, \quad \forall (\ell, j) \in \mathbb{Z}^v \times \mathbb{Z} \quad (28)$$

for some  $\gamma \in (0, 1)$  small enough and  $\tau < v$ . By assuming the condition eq. (28), the operator  $D_u \mathcal{F}(0, 0)$  is formally invertible, that is,

$$D_u \mathcal{F}(0, 0)^{-1}[h] = \sum_{\ell, j} \frac{\widehat{h}(\ell, j)}{i(\omega \cdot \ell + \Lambda(j))} e^{i\ell \cdot \varphi} e^{ijx}$$

but it maps  $H^s \rightarrow H^{s-\tau}$ . Because of this *loss of derivatives*, the standard implicit function theorem cannot be applied. In order to overcome the small divisor's difficulty one then needs to perform a very fast convergent iterative scheme, in order to compensate for the loss of derivatives due to the small divisors. A possible approach is to implement a Nash-Moser scheme (cf. the monograph Nirenberg (2001)), which is essentially a generalization of the Newton tangent method. More precisely, one constructs approximate solutions iteratively, by setting

$$u_0 := 0, \quad u_{n+1} = u_n - \prod_n [D_u \mathcal{F}(u_n, \varepsilon)]^{-1} \mathcal{F}(u_n, \varepsilon)$$

where  $\prod_n : L^2(\mathbb{T}^v \times \mathbb{T}) \rightarrow C^\infty(\mathbb{T}^v \times \mathbb{T})$  is a suitable smoothing operator (for instance the truncation of the Fourier series). This scheme has the advantage to be quadratic in the sense that  $u_{n+1} - u_n$  is controlled essentially by  $(u_n - u_{n-1})^2$  and hence, if  $u_1 - u_0$  is small enough; the scheme converges super exponentially fast. On the other hand, the price that one has to pay for this fast convergence, is that one has to invert the



linearized operator  $D_u \mathcal{F}(u, \varepsilon)$  in a full neighborhood of the origin  $u = 0$  and not only at  $u = 0$  as in the classical implicit function theorem. By eq. (27), the linearized operator  $\mathcal{L} \equiv \mathcal{L}(u)$  has the form

$$\begin{aligned} \mathcal{L} &:= \omega \cdot \partial_\varphi + i\Lambda(D) + \varepsilon \mathcal{R}, \quad \mathcal{R}[h] := D_u Q(u)[h], \\ \mathcal{R} : H^{s+\sigma} &\rightarrow H^s \end{aligned} \quad (29)$$

This operator is a small perturbation  $\varepsilon \mathcal{R}$  of a diagonal operator with small divisors. The problem of the loss of derivatives implies that there is no hope to invert  $\mathcal{L}$  by Neumann series. Moreover the fact that  $\mathcal{R}$  is unbounded creates a very strong perturbative effect to the unperturbed spectrum of  $i\Lambda(D)$ . Actually, the closer  $\sigma = \text{order}(\mathcal{R})$  is to  $m = \text{order}(\Lambda(D))$ , the stronger is the perturbative effect of  $\varepsilon \mathcal{R}$ . Roughly speaking there are two opposite phenomena: (1) on the one hand the perturbation  $\varepsilon \mathcal{R}$  should be negligible, since there is a small parameter  $\varepsilon$ . (2) On the other hand, especially in the fully nonlinear case  $\sigma = m$ , the remainder  $\mathcal{R}$  contains the same number of derivatives as the diagonal part  $\Lambda(D)$  and hence, it cannot be really considered as a perturbation.

This is the reason why the KAM theory has been first developed for PDEs with bounded perturbations or *weakly unbounded* perturbations ( $\sigma \ll m$ ). We shall explain how to overcome these difficulties in the case of the Water Waves equations.

Existence of small amplitude time periodic pure gravity standing wave solutions (periodic separately in time and space) has been proved by Iooss, Plotnikov, Toland in Iooss et al. (2005), Iooss and Plotnikov (2005) and in Plotnikov and Toland (2001) in finite depth (see also Alazard and Baldi (2015) for a more recent extension to the gravity-capillary case). We also recall that existence of periodic, small amplitude 2-d traveling gravity water wave solutions dates back to Levi Civita (1925), see also the other references in section “[Introduction](#)” (standing waves are not traveling because they are even in space, see eq. (31)). The case of periodic standing waves is much more difficult than periodic traveling waves since they are not stationary in a moving frame. This implies that the small divisor’s problem appears in this case. Existence of small amplitude

3-d traveling gravity-capillary water wave solutions with space periodic boundary conditions has been proved in Craig and Nicholls (2000) (it is not a small divisor problem) and in Iooss and Plotnikov (2009)-Iooss and Plotnikov (2011) in the case of zero surface tension (in such a case it is a small divisor problem).

The pioneering idea to prove the existence of periodic and quasi-periodic solutions for fully nonlinear PDEs has been introduced in Iooss et al. (2005), for proving the existence of periodic standing Water Waves. It consists in developing a normal form method by using pseudo-differential calculus, which allows to invert the linearized equations at any approximate solution of a Nash-Moser iteration. This method has been extended also to the harder problem of constructing quasi-periodic solutions, starting from 2014, see Baldi et al. (2014, 2016). For the Water Waves, existence of quasi-periodic standing wave solutions has been proved both in the gravity-capillary case Berti and Montalto (2020) and in the pure gravity case Baldi et al. (2018). We still refer to Berti (2016) for further references.

In the following, we shall describe in detail the results concerning the existence of quasi-periodic standing wave solutions obtained in Berti and Montalto (2020) (gravity capillary Water Waves in infinite depth) and in Baldi et al. (2018) (pure gravity Water Waves in finite depth).

The water-waves system eqs. (21)–(8) exhibits several symmetries. First of all, the mass  $\int_{\mathbb{T}} \eta dx$  is a constant of motion for the water waves eq. (21). Moreover

$$\partial_t \int_{\mathbb{T}} \psi dx = - \int_{\mathbb{T}} \eta dx - \int_{\mathbb{T}} \nabla_\eta K dx = - \int_{\mathbb{T}} \eta dx$$

because  $\int_{\mathbb{T}} \nabla_\eta K dx = 0$ . This follows because  $\mathbb{R} \ni c \mapsto K(c + \eta, \psi)$  is constant (the bottom of the ocean is at  $-\infty$ ) and so  $0 = d_\eta K(\eta, \psi)[1] = (\nabla_\eta K, 1)_{L^2(\mathbb{T})}$ . As a consequence the subspace

$$\int_{\mathbb{T}} \eta dx = \int_{\mathbb{T}} \psi dx = 0 \quad (30)$$

is invariant under the evolution of eq. (21), and we shall restrict to solutions satisfying eq. (30).

In addition, the subspace of functions which are even in  $x$ ,

$$\eta(x) = \eta(-x), \quad \psi(x) = \psi(-x), \quad (31)$$

is invariant under eq. (21). Thanks to this property and eq. (30), we shall restrict  $(\eta, \psi)$  to the phase space of  $2\pi$ -periodic functions, even in  $x$ , which admits the Fourier expansion

$$\begin{aligned} \eta(x) &= \sum_{j \geq 1} \eta_j \cos(jx), \quad \psi(x) \\ &= \sum_{j \geq 1} \psi_j \cos(jx). \end{aligned} \quad (32)$$

In this case also the velocity potential  $\Phi(x, y)$  is even and  $2\pi$ -periodic in  $x$  and so the  $x$ -component of the velocity field  $v = (\Phi_x, \Phi_y)$  vanishes at  $x = k\pi, \forall k \in \mathbb{Z}$ . Hence there is no flux of fluid through the lines  $x = k\pi, k \in \mathbb{Z}$ , and a solution of eq. (21) satisfying eq. (32) describes the motion of a liquid confined between two walls.

Another important symmetry of the capillary Water Waves system is reversibility, namely the eqs. (21)–(8) are reversible with respect to the involution  $\rho : (\eta, \psi) \mapsto (\eta, -\psi)$ , or, equivalently, the Hamiltonian is even in  $\psi$ :

$$\begin{aligned} H \circ \rho &= H, \quad H(\eta, \psi) = H(\eta, -\psi), \quad \rho \\ &: (\eta, \psi) \mapsto (\eta, -\psi). \end{aligned} \quad (33)$$

As a consequence it is natural to look for solutions of eq. (21) satisfying

$$\begin{aligned} u(-t) &= \rho u(t), \quad i.e. \eta(-t, x) \\ &= \eta(t, x), \psi(-t, x) \\ &= -\psi(t, x), \forall t, x \in \mathbb{R}, \end{aligned} \quad (34)$$

namely,  $\eta$  is even in time and  $\psi$  is odd in time. Solutions of the Water Waves eqs. (21), satisfying (32) and (34) are called *standing Water Waves*.

### The KAM for Gravity Capillary Water Waves

Let us present rigorously the KAM result obtained in Berti and Montalto (2020) on gravity capillary Water Waves in infinite depth. As already said we look for small amplitude quasi-periodic solutions of eq. (21). It is therefore of main importance to analyze the dynamics of the linearized system

eq. (21), at the equilibrium  $(\eta, \psi) = (0, 0)$  (flat ocean and fluid at rest), namely,

$$\begin{cases} \partial_t \eta = G(0)\psi, \\ \partial_t \psi + \eta = \kappa \eta_{xx} \end{cases} \quad (35)$$

where  $G(0) = |D|$  (infinite depth) is the Dirichlet-Neumann operator for the flat surface  $\eta = 0$ , namely

$$\begin{aligned} |D| \cos(jx) &= |j| \cos(jx), \quad |D| \sin(jx) \\ &= |j| \sin(jx), \forall j \in \mathbb{Z}. \end{aligned}$$

In compact Hamiltonian form, the system eq. (51) reads

$$\partial_t u = J \Omega u, \quad \Omega := \begin{pmatrix} 1 - \kappa \partial_{xx} & 0 \\ 0 & G(0) \end{pmatrix}, \quad (36)$$

which is the Hamiltonian system generated by the quadratic Hamiltonian (see eq. (9))

$$\begin{aligned} H_L &: \frac{1}{2} (u, \Omega u)_{L^2(\mathbb{T}_x)} \\ &= \frac{1}{2} (\psi, G(0)\psi)_{L^2(\mathbb{T}_x)} \\ &\quad + \frac{1}{2} \int_{\mathbb{T}} (\eta^2 + \kappa \eta_x^2) dx. \end{aligned} \quad (37)$$

The standing wave solutions of the linear system eq. (51), that is, eq. (52), are

$$\begin{aligned} \eta(t, x) &= \sum_{j \geq 1} a_j \cos(\omega_j t) \cos(jx), \\ \psi(t, x) &= -\sum_{j \geq 1} a_j j^{-1} \omega_j \sin(\omega_j t) \cos(jx), \end{aligned} \quad (38)$$

$a_j \in \mathbb{R}$ , with linear frequencies of oscillations

$$\omega_j := \omega_j(\kappa) := \sqrt{j(1 + \kappa j^2)}, j \geq 1. \quad (39)$$

The main aim is to prove that most of the standing wave solutions eq. (54) of the linear system eq. (51) can be continued to standing wave solutions of the nonlinear water-waves Hamiltonian system eq. (21) for most values of the surface tension parameter  $\kappa \in [\kappa_1, \kappa_2]$ . More precisely, fix an arbitrary finite subset  $\mathbb{S}^+ \subset \mathbb{N}^+ := \{1, 2, \dots\}$

(called “tangential sites”) and consider the linear standing wave solutions (of eq. (51))

$$\begin{aligned}\eta(t, x) &= \sum_{j \in \mathbb{S}^+} \sqrt{\xi_j} \cos(\omega_j t) \cos(jx), \\ \psi(t, x) &= - \sum_{j \in \mathbb{S}^+} \sqrt{\xi_j} j^{-1} \omega_j \sin(\omega_j t) \cos(jx), \\ \xi_j &> 0,\end{aligned}\quad (40)$$

which are Fourier supported in  $\mathbb{S}^+$ . In Theorem 4.2 below we prove the existence of quasi-periodic solutions  $u(\tilde{\omega}t, x) = (\eta, \psi)(\tilde{\omega}t, x)$  of eq. (21), with frequency  $\tilde{\omega} := (\tilde{\omega}_j)_{j \in \mathbb{S}^+}$ ,  $\tilde{\omega}_j \simeq \omega_j$ , close to the solutions eq. (56) of eq. (51), for most values of the surface tension parameter  $\kappa \in [\kappa_1, \kappa_2]$ .

Let  $v := |\mathbb{S}^+|$  denote the cardinality of  $\mathbb{S}^+$ . The function  $u(\varphi, x) = (\eta, \psi)(\varphi, x)$ ,  $\varphi \in \mathbb{T}^v$ , belongs to the Sobolev spaces of  $(2\pi)^{v+1}$ -periodic real functions

$$\begin{aligned}H^s(\mathbb{T}^{v+1}, \mathbb{R}^2) &:= \{u = (\eta, \psi) : \eta, \psi \in H^s\} \\ H^s &:= H^s(\mathbb{T}^{v+1}, \mathbb{R}) = \left\{ f = \sum_{(\ell, j) \in \mathbb{Z}^{v+1}} \hat{f}_{\ell, j} e^{i(\ell \cdot \varphi + jx)} : \right. \\ &\quad \left. \|f\|_s^2 := \sum_{(\ell, j) \in \mathbb{Z}^{v+1}} |\hat{f}_{\ell, j}|^2 \langle \ell, j \rangle^{2s} < +\infty \right\}\end{aligned}\quad (41)$$

where  $\langle \ell, j \rangle := \max\{1, |\ell|, |j|\}$  with  $|\ell| := \max_{i=1, \dots, v} |\ell_i|$ . For

$$s \geq s_0 := \left\lceil \frac{v+1}{2} \right\rceil + 1 \in \mathbb{N} \quad (42)$$

the Sobolev spaces  $H^s \subset L^\infty(\mathbb{T}^{v+1})$  is algebra with respect to the product of functions.

**Theorem 4.1 (KAM for capillary-gravity Water Waves (Berti and Montalto 2020))** *For every choice of finitely many tangential sites  $\mathbb{S}^+ \subset \mathbb{N}^+$ , there exists  $\bar{s} > s_0$ ,  $\varepsilon_0 \in (0, 1)$  such that for every  $|\xi| \leq \varepsilon_0^2$ ,  $\xi := (\xi_j)_{j \in \mathbb{S}^+}$ ,  $\xi_j > 0$  for any  $j \in \mathbb{S}^+$ , there exists a Borel set  $\mathcal{G} \subset [\kappa_1, \kappa_2]$  with asymptotically full measure as  $\xi \rightarrow 0$ , that is,*

$$\lim_{\xi \rightarrow 0} |\mathcal{G}| = \kappa_2 - \kappa_1,$$

*such that, for any surface tension coefficient  $\kappa \in \mathcal{G}$ , the capillary-gravity system eq. (21) has a time quasi-periodic standing wave solution.*

$$u(\tilde{\omega}t, x) = (\eta(\tilde{\omega}t, x), \psi(\tilde{\omega}t, x)).$$

*with Sobolev regularity  $(\eta, \psi) \in H^{\bar{s}}(\mathbb{T}^v \times \mathbb{T}, \mathbb{R}^2)$ , of the form*

$$\begin{aligned}\eta(\tilde{\omega}t, x) &= \sum_{j \in \mathbb{S}^+} \sqrt{\xi_j} \cos(\tilde{\omega}_j t) \cos(jx) \\ &\quad + r_1(\tilde{\omega}t, x), \\ \psi(\tilde{\omega}t, x) &= - \sum_{j \in \mathbb{S}^+} \sqrt{\xi_j} j^{-1} \omega_j \sin(\tilde{\omega}_j t) \cos(jx) \\ &\quad + r_2(\tilde{\omega}t, x)\end{aligned}\quad (43)$$

*with a diophantine frequency vector  $\tilde{\omega} := \tilde{\omega}(\kappa, \xi) \in \mathbb{R}^v$  satisfying  $\tilde{\omega}_j - \omega_j(\kappa) \rightarrow 0$ ,  $j \in \mathbb{S}^+$ , as  $\xi \rightarrow 0$ , and the functions  $r_1(\varphi, x)$ ,  $r_2(\varphi, x)$  are  $o(\sqrt{|\xi|})$ -small in  $H^{\bar{s}}(\mathbb{T}^v \times \mathbb{T}, \mathbb{R})$ , that is  $\|r_j\|_{\bar{s}} / \sqrt{|\xi|}$  tends to 0 as  $|\xi| \rightarrow 0$  for  $j = 1, 2$ . In addition these quasi-periodic solutions are linearly stable.*

#### Some Ideas of the Proof

Let us present more in details some key ideas of the proof of Theorem 4.2. As we explained at the beginning of section “KAM Results for Water Waves,” see eqs. (27)–(29), the general idea is to prove the existence of quasi-periodic solutions in Theorem 4.2, by using a Nash-Moser iteration. In the case of gravity capillary Water Waves, the problem of inverting the linearized operator is essentially equivalent to invert (for most values of the parameters  $\kappa \in [\kappa_1, \kappa_2]$ ) a linear pseudo-differential operator of the form

$$\begin{aligned}\mathcal{L} &:= \omega \cdot \partial_\varphi + i(1 + \varepsilon a(\varphi, x))T(D) + \varepsilon b(\varphi, x)\partial_x \\ &\quad + \varepsilon c(\varphi, x)\mathcal{H}|D|^{\frac{1}{2}} + \varepsilon \text{Op}(r(\varphi, x, \xi)), \\ T(D) &:= \sqrt{|D|(1 - \kappa \partial_{xx})}\end{aligned}\quad (44)$$

where  $a, b, c \in C^\infty(\mathbb{T}^{v+1})$ ,  $r \in \mathcal{S}^0$  is a symbol of order 0 (recall the definition 3.1) and  $\mathcal{H}$  is the Hilbert transform, namely the Fourier multiplier

$$\mathcal{H}u(x) := \sum_{\xi \in \mathbb{Z}} i \operatorname{sign}(\xi) \widehat{u}(\xi) e^{ix\xi}.$$

In order to invert  $\mathcal{L}$ , one constructs by normal form techniques, combined with pseudo-differential operator, a transformation  $\Phi$  which completely diagonalizes the operator  $\mathcal{L}$ . This normal form construction is split in two parts:

- **Reduction in decreasing orders.** One constructs a bounded, invertible transformation  $\mathcal{P} : H^s \rightarrow H^s$ ,  $s \gg 0$  such that

$$\mathcal{L}_0 : \mathcal{P}^{-1} \mathcal{L} \mathcal{P} = \omega \cdot \partial_\varphi + i\mathcal{D}_0 + \varepsilon \mathcal{R}_0(\varphi) \quad (45)$$

where  $\mathcal{D}_0$  is a diagonal operator  $\mathcal{D}_0 = \operatorname{diag}_{j \geq 1} \mu_j$  with eigenvalues  $\mu_j \in \mathbb{R}$ ,  $j \geq 1$  and  $\mathcal{R}_0 \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T})))$  is a bounded linear operator for any  $s \gg 0$  large enough. Here  $\mathcal{B}(H^s(\mathbb{T}))$  denotes the space of bounded linear operators on  $H^s$ . The eigenvalues  $\mu_j$  have the asymptotic expansion

$$\begin{aligned} \mu_j &= (1 + \varepsilon m_1) \sqrt{j(1 + \kappa j^2)} + \varepsilon m_2 j^{\frac{1}{2}} \\ \text{for some constants } m_1, m_2 &\in \mathbb{R}, m_1, m_2 \\ &= O(\varepsilon). \end{aligned} \quad (46)$$

This pseudo-differential normal form procedure *reduces the order of the perturbation but not the size*. Indeed the remainder  $\varepsilon \mathcal{R}_0$  is a bounded linear operator  $H^s \rightarrow H^s$ , but its size is still  $\sim \varepsilon$  (as the size of the original perturbation in eq. (44)). The normal form transformation  $\mathcal{P}$  is constructed in several steps, by means of flows generated by suitable pseudo-differential operators. For instance, in order to normalize the highest order term  $(1 + \varepsilon a(\varphi, x)) T(D)$  in eq. (44), one considers composition operators induced by diffeomorphism of the torus

$$\begin{aligned} x &\mapsto x + \varepsilon \beta(\psi, x), \beta \in C^\infty(\mathbb{T}^v \times \mathbb{T}) \\ \mathcal{B} : h(\varphi, x) &\mapsto h(\varphi, x + \varepsilon \beta(\varphi, x)) \end{aligned}$$

By choosing  $\beta(\varphi, x)$  and  $\lambda(\varphi)$  (independent of  $x$ ) in such a way that

$$(1 + \varepsilon \beta_x(\varphi, x))^{\frac{3}{2}} (1 + \varepsilon a(\varphi, x)) = 1 + \varepsilon \lambda(\varphi)$$

one obtains that

$$\mathcal{B}^{-1} \mathcal{L} \mathcal{B} = \omega \cdot \partial_\varphi + (1 + \varepsilon \lambda(\varphi)) T(D) + \dots$$

where  $\dots$  denote the terms which are of order smaller than one. To remove the dependence on  $\varphi$  from the highest order term, one performs a similar transformation by reparametrizing the time.

- **The KAM reducibility scheme.** Once one deals with  $\mathcal{L}_0$  in eq. (45) one performs a KAM-reducibility scheme, in order to reduce *quadratically* the size of the remainder  $\varepsilon \mathcal{R}_0$ . Along this iterative scheme, one imposes second Melnikov conditions of the form

$$|\omega \cdot \ell + \mu_j - \mu_{j'}| \geq \frac{\gamma |j^{\frac{3}{2}} - j'^{\frac{3}{2}}|}{\langle \ell \rangle^\tau}, \quad \forall (\ell, j, j') \neq (0, j, j) \quad (47)$$

for  $\gamma \in (0, 1)$  small enough. Let us describe roughly the first step of this iterative procedure. The aim is to diagonalize  $\mathcal{L}_0$  up to order  $O(\varepsilon^2)$ . We then consider a map  $\varphi \mapsto \Psi(\varphi)$  which has to be chosen appropriately and compute

$$\begin{aligned} e^{-\varepsilon \Psi(\varphi)} \mathcal{L}_0 e^{\varepsilon \Psi(\varphi)} &= \omega \cdot \partial_\varphi + \mathcal{D}_0 + \varepsilon (\omega \cdot \partial_\varphi \Psi(\varphi) \\ &\quad + i[\mathcal{D}_0, \Psi(\varphi)] + \mathcal{R}_0(\varphi)) + O(\varepsilon^2). \end{aligned} \quad (48)$$

If we have a smooth map

$$\mathbb{T}^v \rightarrow \mathcal{B}(L^2(\mathbb{T})), \quad \varphi \mapsto \mathcal{R}(\varphi)$$

we can write

$$\mathcal{R}(\varphi) = \sum_{\ell \in \mathbb{Z}^v} \widehat{\mathcal{R}}(\ell) e^{i\ell \cdot \varphi}, \quad \widehat{\mathcal{R}}(\ell) \in \mathcal{B}(L^2(\mathbb{T})).$$

Since all the operators involved are invariant on the space of even functions, we can represent  $\widehat{\mathcal{R}}(\ell)$  with respect to the basis  $\cos(jx)$ ,  $j \geq 1$  by obtaining a matrix representation  $\widehat{\mathcal{R}}(\ell)_j^{j'}$ ,  $j, j' \in \mathbb{N}^+$ .

By using this formalism, we then solve the homological equation

$$\omega \cdot \partial_\varphi \Psi(\varphi) + i[\mathcal{D}, \Psi(\varphi)] + \mathcal{R}_0(\varphi) = Z$$

where  $Z = \text{diag}_{j \geq 1} z_j$ ,  $z_j := \widehat{\mathcal{R}}(0)_j^j$ ,  $j \geq 1$  and by defining the matrix representation of  $\Psi$  in the following way:

$$\widehat{\Psi}(\ell)_j^{j'} := \begin{cases} -\frac{\widehat{\mathcal{R}}_0(\ell)_j^{j'}}{i(\omega \cdot \ell + \mu_j - \mu_{j'})} & \forall (\ell, j, j') \neq (0, j, j) \\ 0 & \text{otherwise.} \end{cases} \quad (49)$$

By the nonresonance condition eq. (47) and using that  $\mathcal{R}_0 \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T})))$ ,  $\forall s \gg 0$ , one gets that also  $\Psi \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T})))$  and then the remainder of order  $O(\varepsilon^2)$  in eq. (48) is still in  $C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T})))$ . By iterating this procedure, one then gets a complete diagonalization of the operator  $\mathcal{L}_0$  in eq. (45). More precisely one constructs a bounded and invertible transformation:  $\Phi_\infty : H^s(\mathbb{T}^v \times \mathbb{T}) \rightarrow H^s(\mathbb{T}^v \times \mathbb{T})$ ,  $\forall s \gg 0$  such that

$$\Phi_\infty^{-1} \mathcal{L}_0 \Phi_\infty = \omega \cdot \partial_\varphi + i\mathcal{D}_\infty, \quad \mathcal{D}_\infty := \text{diag}_{j \geq 1} \mu_j^\infty$$

and  $\mu_j^\infty \in \mathbb{R}$ ,

$$\mu_j^\infty = (1 + \varepsilon m_1) \sqrt{j(1 + \kappa j^2)} + \varepsilon m_2 j^{\frac{1}{2}} + r_j^\infty, \\ \sup_{j \geq 1} r_j^\infty = O(\varepsilon)$$

(recall eq. (46)). Note that the reversible structure eqs. (33 and 34) is crucial in the normal form procedure, since it allows to *average out* some secular and dissipative terms.

A final comment concerns the nonresonance conditions eq. (47), which have to be verified along the KAM procedure. They can be verified by degenerating the KAM theory (explained below) exploiting that the unperturbed linear frequencies  $\kappa \mapsto \omega_j(\kappa)$  are *analytic*, simple, grow asymptotically as  $j^{3/2}$  and are *nondegenerate* in the sense of Bambusi-Berti-Magistrelli (2011).

For such values of  $\kappa \in [\kappa_1, \kappa_2]$ , the solutions eq. (56) of the linear eq. (51) are already sufficiently good approximate quasi-periodic solutions of the nonlinear Water Waves system eq. (21).

### The KAM for Pure Gravity Water Waves in Finite Depth

We now describe the KAM result obtained in Baldi et al. (2018) for the pure gravity Water Waves. One looks for small amplitude solutions of eq. (21), with  $\kappa = 0$ , namely the system

$$\begin{cases} \partial_t \eta = G(\eta) \psi, \\ \partial_t \psi + \eta + \frac{1}{2} \psi_x^2 - \frac{1}{2} \frac{(G(\eta) \psi + \eta_x \psi_x)^2}{1 + \eta_x^2} = 0 \end{cases} \quad (50)$$

The linearized dynamic at the equilibrium  $(\eta, \psi) = (0, 0)$  is given by

$$\begin{cases} \partial_t \eta = G(0) \psi \\ \partial_t \psi = -\eta \end{cases} \quad (51)$$

where  $G(0) = |D| \tanh(hD)$  is the Dirichlet-Neumann operator at the flat surface  $\eta = 0$ . In the compact Hamiltonian form as in eq. (8), system eq. (51) reads

$$\partial_t u = J \Omega u, \quad \Omega : \begin{pmatrix} 1 & 0 \\ 0 & G(0) \end{pmatrix}, \quad (52)$$

which is the Hamiltonian system generated by the quadratic Hamiltonian (see eq. (9))

$$H_L := \frac{1}{2} (u, \Omega u)_{L^2} = \frac{1}{2} \int_{\mathbb{T}} \psi G(0) \psi dx + \frac{1}{2} \int_{\mathbb{T}} \eta^2 dx. \quad (53)$$

The solutions of the linear system eq. (51), that is, eq. (52), even in  $x$ , satisfying eq. (34) are

$$\begin{aligned} \eta(t, x) &= \sum_{j \geq 1} \xi_j \cos(\omega_j t) \cos(jx), \quad \psi(t, x) \\ &= -\sum_{j \geq 1} \xi_j \omega_j^{-1} \sin(\omega_j t) \cos(jx) \end{aligned} \quad (54)$$

where

$$\omega_j := \omega_j(\mathbf{h}) := \sqrt{j \tanh(\mathbf{h}j)}, j \geq 1. \quad (55)$$

Note that, since  $j \mapsto j \tanh(\mathbf{h}j)$  is monotone increasing, all the linear frequencies are simple. One shows that most solutions eq. (54) of the linear system eq. (51) can be continued to solutions of the nonlinear Water Waves eq. (50) for most values of the depth parameter  $\mathbf{h} \in [\mathbf{h}_1, \mathbf{h}_2]$ . More precisely one looks for quasi-periodic solutions  $u(\tilde{\omega}t) = (\eta, \psi)(\tilde{\omega}t)$  with frequency  $\tilde{\omega} \in \mathbb{R}^v$  (to be determined), close to solutions eq. (54) of eq. (51), in the Sobolev spaces of functions  $H^s(\mathbb{T}^v \times \mathbb{T})$ . Fix an arbitrary finite subset  $\mathbb{S}_+ \subset \mathbb{N}^+ := \{1, 2, \dots\}$  (tangential sites) and consider the solutions of the linear eq. (51)

$$\begin{aligned} \eta(t, x) &= \sum_{j \in \mathbb{S}_+} \xi_j \cos(\omega_j(\mathbf{h})t) \cos(jx), \psi(t, x) \\ &= - \sum_{j \in \mathbb{S}_+} \frac{\xi_j}{\omega_j(\mathbf{h})} \sin(\omega_j(\mathbf{h})t) \cos(jx), \xi_j \\ &> 0, \end{aligned} \quad (56)$$

which are Fourier supported on  $\mathbb{S}_+$ . We denote by  $v := |\mathbb{S}_+|$  the cardinality of  $\mathbb{S}_+$ .

**Theorem 4.2 (KAM for gravity Water Waves in finite depth (Baldi et al. (2018))** *For every choice of the tangential sites  $\mathbb{S}_+ \subset \mathbb{N} \setminus \{0\}$ , there exists  $\bar{s} > \frac{|\mathbb{S}_+|+1}{2}$ ,  $\varepsilon_0 \in (0, 1)$  such that for every vector  $\vec{\xi} := (\xi_j)_{j \in \mathbb{S}_+}$ , with  $\xi_j > 0$  for all  $j \in \mathbb{S}_+$  and  $|\vec{\xi}| \leq \varepsilon_0$ , there exists a Cantor-like set  $\mathcal{G} \subset [\mathbf{h}_1, \mathbf{h}_2]$  with asymptotically full measure as  $\vec{\xi} \rightarrow 0$ , that is,*

$$\lim_{\vec{\xi} \rightarrow 0} |\mathcal{G}| = \mathbf{h}_2 - \mathbf{h}_1,$$

*such that, for any  $\mathbf{h} \in \mathcal{G}$ , the gravity Water Waves system eq. (50) has a time quasi-periodic solution*

*$u(\tilde{\omega}t, x) = (\eta(\tilde{\omega}t, x), \psi(\tilde{\omega}t, x))$ , with Sobolev regularity  $(\eta, \psi) \in H^{\bar{s}}(\mathbb{T}^v \times \mathbb{T}, \mathbb{R}^2)$ , with a Diophantine frequency vector  $\tilde{\omega} := \tilde{\omega}(\mathbf{h}, \vec{\xi}) := (\tilde{\omega}_j)_{j \in \mathbb{S}_+} \in \mathbb{R}^v$ , of the form*

$$\begin{aligned} \eta(\tilde{\omega}t, x) &= \sum_{j \in \mathbb{S}_+} \xi_j \cos(\tilde{\omega}_j t) \cos(jx) + r_1(\tilde{\omega}t, x), \\ \psi(\tilde{\omega}t, x) &= - \sum_{j \in \mathbb{S}_+} \frac{\xi_j}{\omega_j(\mathbf{h})} \sin(\tilde{\omega}_j t) \cos(jx) \\ &\quad + r_2(\tilde{\omega}t, x) \end{aligned} \quad (57)$$

*with  $\tilde{\omega}(\mathbf{h}, \vec{\xi}) \rightarrow \tilde{\omega}(\mathbf{h}) := (\omega_j(\mathbf{h}))_{j \in \mathbb{S}_+}$  as  $\vec{\xi} \rightarrow 0$ , and the functions  $r_1(\varphi, x)$ ,  $r_2(\varphi, x)$  are  $o(|\vec{\xi}|)$ -small in  $H^{\bar{s}}(\mathbb{T}^v \times \mathbb{T}, \mathbb{R})$ , that is,  $\|r_i\|_{\bar{s}} / |\vec{\xi}| \rightarrow 0$  as  $|\vec{\xi}| \rightarrow 0$  for  $i = 1, 2$ . The solution  $(\eta(\tilde{\omega}t, x), \psi(\tilde{\omega}t, x))$  is even in  $x$ ,  $\eta$  is even in  $t$  and  $\psi$  is odd in  $t$ . In addition these quasi-periodic solutions are linearly stable.*

#### Ideas of the Proof

We first outline the deep differences between the pure gravity Water Waves system and the gravity capillary Water Waves system, discussed in the previous section. The general strategy implemented to prove Theorem 4.2 is the same as the one explained in section “Some Ideas of the Proof.” Essentially at each step of the Nash-Moser scheme one has to invert a linear operator of the form

$$\begin{aligned} \mathcal{L} := & \omega \cdot \partial_\varphi + i(1 + \varepsilon a(\varphi, x)) \Lambda(D) + \varepsilon V(\varphi, x) \partial_x \\ & + \varepsilon \text{Op}(r_0(\varphi, x, \xi)), \\ \Lambda(D) := & (|D| \tanh(\mathbf{h}D))^{\frac{1}{2}}, \\ & a, V \in C^\infty(\mathbb{T}^v \times \mathbb{T}), \quad r_0 \in S^0 \end{aligned} \quad (58)$$

where we recall the definition 3.1. The main additional difficulties are the following ones:



- (i) The linearized operator  $\mathcal{L}$  in eq. (58) is a *singular perturbation* of the unperturbed ( $\varepsilon = 0$ ) linear operator  $\omega \cdot \partial_\varphi + i\Lambda(D)$ . Indeed, the transport term  $\varepsilon V \partial_x$  is a small perturbation of order one of the unperturbed operator  $[|D| \tanh(hD)]^{\frac{1}{2}}$  of order 1/2. On the other hand, the gravity capillary vector field is quasi-linear and contains derivatives of the same order as the linearized vector field at the origin, which is  $\sim |D|^{\frac{3}{2}}$ . This difference, which is well known in the Water Waves literature, requires a very different analysis of the linearized operator with respect to the gravity capillary case in Berti and Montalto (2020).
- (ii) The linear frequencies  $\omega_j$  in eq. (55) of the pure gravity Water Waves grow like  $\sim j^{\frac{1}{2}}$  as  $j \rightarrow +\infty$ , while, in the presence of surface tension  $\kappa$ , the linear frequencies are  $\sqrt{(1 + \kappa j^2)j \tanh(hj)} \sim j^{\frac{3}{2}}$ . This makes a substantial difference for the development of the KAM theory. In presence of a sub-linear growth of the linear frequencies  $\sim j^\alpha$ ,  $\alpha < 1$ , one may impose only very weak second order Melnikov non-resonance conditions, which loose also space (and not only time) derivatives along the KAM reducibility scheme.
- (iii) The linear frequencies eq. (55) vary with  $h$  only by exponentially small quantities: they admit the asymptotic expansion

$$\begin{aligned} \sqrt{j \tanh(hj)} &= \sqrt{j} + r(j, h) \text{ where } |\partial_h^k r(j, h)| \\ &\leq C_k e^{-hj} \forall k \in \mathbb{N}, \forall j \geq 1, \end{aligned} \quad (59)$$

uniformly in  $h \in [h_1, h_2]$ , where the constant  $C_k$  depends only on  $k$  and  $h_1$ . Nevertheless we shall be able, extending the degenerate KAM theory approach in Bambusi et al. (2011), Berti and Montalto (2020), to use the finite depth parameter  $h$  to impose the required Melnikov nonresonance conditions. On the other hand, for the gravity capillary Water Waves considered in Berti and Montalto (2020), the surface tension parameter  $\kappa$  moves the linear

frequencies  $\sqrt{(1 + \kappa j^2)j \tanh(hj)}$  of polynomial quantities  $O(j^{3/2})$ .

As in the gravity capillary case, the normal form procedure on the operator  $\mathcal{L}$  (in 4.38) is split in two parts that we shall describe below.

- **Reduction in decreasing order.** One constructs a bounded, invertible transformation  $\mathcal{P} : H^s \rightarrow H^s$ ,  $s \gg 0$  such that

$$\mathcal{L}_0 := \mathcal{P}^{-1} \mathcal{L} \mathcal{P} = \omega \cdot \partial_\varphi + i\mathcal{D}_0 + \varepsilon \mathcal{R}_0(\varphi) \quad (60)$$

where  $\mathcal{D}_0$  is a diagonal operator  $\mathcal{D}_0 = \text{diag}_{j \geq 1} \mu_j$  with eigenvalues  $\mu_j \in \mathbb{R}$ ,  $j \geq 1$  and  $\mathcal{R}_0 \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T}), H^{s+M}(\mathbb{T})))$ ,  $M \gg 0$ ,  $s \gg 0$  large enough. The eigenvalues  $\mu_j$  have the asymptotic expansion

$$\begin{aligned} \mu_j &= (1 + \varepsilon m) \sqrt{j \tanh(hj)} + \varepsilon r_j \\ &= O(j^{-\frac{1}{2}}). \end{aligned} \quad (61)$$

The regularization up to an arbitrary smoothing remainder is essential in order to perform a KAM procedure with nonresonance conditions losing space derivatives. This is a feature of the sublinear dispersion  $\sim \sqrt{j}$ , which implies that the differences  $\sqrt{j} - \sqrt{j'}$  accumulate to 0. This does not happen in the gravity capillary case, since  $j^{\frac{3}{2}} - j'^{\frac{3}{2}} \gg 1$  if  $j \neq j'$ . In order to remove the *singular perturbation term*  $\varepsilon V(\varphi, x) \partial_x$ , we strongly use the reversibility of the Water Waves system. Indeed, it turns out that  $V = \text{odd}(\varphi) \text{odd}(x)$ , meaning that

$$\begin{aligned} V(\varphi, x) &= -V(-\varphi, x), \quad \forall (\varphi, x) \in \mathbb{T}^v \times \mathbb{T}, \\ V(\varphi, x) &= -V(\varphi, -x), \quad \forall (\varphi, x) \in \mathbb{T}^v \times \mathbb{T}. \end{aligned}$$

We then consider a change of variables induced by a diffeomorphism of the torus

$$\begin{aligned} x &\mapsto x + \varepsilon \beta(\varphi, x), \quad \beta \in C^\infty(\mathbb{T}^v \times \mathbb{T}) \\ \mathcal{B} &: h(\varphi, x) \mapsto h(\varphi, x + \varepsilon \beta(\varphi, x)) \end{aligned}$$

By choosing  $\beta(\varphi, x)$  in such a way that

$$\omega \cdot \partial_\varphi \beta(\varphi, x) + \varepsilon V(\varphi, x)(1 + \partial_x \beta(\varphi, x)) = 0 \quad (62)$$

one obtains that

$$\mathcal{B}^{-1} \mathcal{L} B = \omega \cdot \partial_\varphi + (1 + \varepsilon a(\varphi, x)) \Lambda(D) + \dots$$

where  $\dots$  denote the terms of order smaller than  $1/2$ . The quasi-periodic transport eq. (62) can be solved by imposing a diophantine condition on  $\omega$ , that is,

$$|\omega \cdot \ell| \geq \frac{\gamma}{\langle \ell \rangle^\tau}, \quad \forall \ell \in \mathbb{Z}^v \setminus \{0\}$$

by a quadratic iterative scheme in which one transforms the operator  $\omega \cdot \partial_\varphi + \varepsilon V(\varphi, x) \partial_x$  to  $\omega \cdot \partial_\varphi$ . The fact that all the averages are zero along this procedure is due to the fact that  $V = \text{odd}(\varphi)$  and this property is preserved along the iteration.

- **The KAM reducibility scheme.** Along the KAM reducibility scheme which diagonalizes  $\mathcal{L}_0$  in eq. (60) one imposes very weak second Melnikov conditions of the form

$$\begin{aligned} |\omega \cdot \ell + \mu_j - \mu_{j'}| &\geq \frac{\gamma}{\langle \ell \rangle^\tau j^\tau j'^\tau}, \quad \forall (\ell, j, j') \\ &\neq (0, j, j) \end{aligned} \quad (63)$$

for  $\gamma \in (0, 1)$  small enough and  $\tau \gg v$  large enough. Let us describe roughly the first step of this iterative procedure. The aim is to diagonalize  $\mathcal{L}_0$  up to order  $O(\varepsilon^2)$ . We then consider a map  $\varphi \mapsto \Psi(\varphi)$  which has to be chosen appropriately and compute

$$\begin{aligned} e^{-\varepsilon \Psi(\varphi)} \mathcal{L}_0 e^{\varepsilon \Psi(\varphi)} &= \omega \cdot \partial_\varphi + \mathcal{D}_0 \\ &+ \varepsilon (\omega \cdot \partial_\varphi \Psi(\varphi) + i[\mathcal{D}, \Psi(\varphi)] + \mathcal{R}_0(\varphi)) \\ &+ O(\varepsilon^2). \end{aligned} \quad (64)$$

One then solves the homological equation

$$\omega \cdot \partial_\varphi \Psi(\varphi) + i[\mathcal{D}, \Psi(\varphi)] + \mathcal{R}_0(\varphi) = \mathcal{Z}$$

where  $\mathcal{Z} = \text{diag}_{j \geq 1} z_j$ ,  $z_j := \widehat{\mathcal{R}}(0)_j^j$ ,  $j \geq 1$  and by defining the matrix representation of  $\Psi$  in the following way:

$$\hat{\Psi}(\ell)_{jj'}^j := \begin{cases} -\frac{\hat{\mathcal{R}}_0(\ell)_j^j}{i(\omega \cdot \ell + \mu_j - \mu_{j'})} & \forall (\ell, j, j') \neq (0, j, j) \\ 0 & \text{otherwise.} \end{cases} \quad (65)$$

By the nonresonance condition eq. (63) and using that  $\mathcal{R}_0 \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T}), H^{s+M}(\mathbb{T})))$ ,  $\forall s \gg 0$ , by choosing the order of regularization  $M$  much bigger than the loss of the small divisor's  $\tau$ , that is,  $M \gg \tau$  one gets that also  $\Psi \in C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T})))$  and then the remainder of order  $O(\varepsilon^2)$  in eq. (48) is still in  $C^\infty(\mathbb{T}^v, \mathcal{B}(H^s(\mathbb{T}), H^{s+M}(\mathbb{T})))$ . By iterating this procedure, one then gets a complete diagonalization of the operator  $\mathcal{L}_0$  in eq. (60). More precisely one constructs a bounded and invertible transformation:  $\Phi_\infty : H^s(\mathbb{T}^v \times \mathbb{T}) \rightarrow H^s(\mathbb{T}^v \times \mathbb{T})$ ,  $\forall s \gg 0$  such that

$$\Phi_\infty^{-1} \mathcal{L}_0 \Phi_\infty = \omega \cdot \partial_\varphi + i \mathcal{D}_\infty, \quad \mathcal{D}_\infty := \text{diag}_{j \geq 1} \mu_j^\infty$$

and  $\mu_j^\infty \in \mathbb{R}$ ,

$$\begin{aligned} \mu_j^\infty &= (1 + \varepsilon m) \sqrt{j \tanh(hj)} + r_j^\infty, \quad r_j^\infty \\ &= O\left(\varepsilon j^{-\frac{1}{2}}\right). \end{aligned}$$

### Quasi-periodic Traveling Water Waves

The results obtained in Berti and Montalto (2020) and in Baldi et al. (2018) have been generalized to quasi-periodic traveling Water Waves in Berti, Franzoi, Maspero (2020b). In this paper, the authors proved the existence of quasi-periodic traveling Water Waves with constant vorticity. The main technical issue is that in order to verify the nonresonance conditions along the KAM procedure, they use the *conservation of momentum*, which allows to *cut away* some resonant Fourier indices. Eventually, we mention also Feola and

Giuliani (2020) in which the authors deal with pure gravity Water Waves in infinite depth. Hence the major difficulty is that there are no *external parameters* to be used in order to impose the non-resonance conditions required. They then perform a careful bifurcation analysis in order to select initial data which they use in order to avoid resonances along the KAM procedure.

## Longtime Existence for Periodic Water Waves

A very important problem in the theory of Water Waves (and also in general of partial differential equations) is to analyze the long time behavior of solutions and possibly to investigate whether or not there are global in time solutions. When the horizontal variable  $x$  is in  $\mathbb{R}^d$ , for sufficiently small and spatially localized initial data, it is possible to construct global-in-time solutions exploiting the dispersive properties of the flow. Results for the 2d case have been proved in Wu (2009), Ionescu and Pusateri (2015, 2016, 2018a), Alazard and Delort (2015a, b), Ifrim and Tataru (2016), Wang (2018) and in Germain et al. (2012), Wu (2011) for the 3d case. We refer again to Ionescu and Pusateri (2018b) and Deng et al. (2017), for a more detailed presentation of these results.

On the other hand, in the case where the horizontal variable  $x$  is in  $\mathbb{T}^d$ , there are no dispersive effects that can be used and up to now, there are no

global in time existence results for the periodic Water Waves equations. A partial substitute (in the periodic setup) is to prove *long time existence results*, which means to show that the solutions can be extended smoothly beyond the time of existence predicted by the local existence theory. In the case of small data of size  $\varepsilon$ , this means extending solutions for times longer than  $O(\varepsilon^{-1})$ . The main tool to prove such results is normal form theory. In order to explain roughly the idea, let us consider a generic PDE as in eq. (25), namely

$$\partial_t u + i\Lambda(D)u + Q(u) = 0, \quad x \in \mathbb{T}^d \quad (66)$$

where  $\Lambda(D)$  is a suitable Fourier multiplier of order  $m$  and  $Q(u)$  is a quadratic form in  $u$

$$Q(u) := \sum_{\xi_1, \xi_2} q(\xi_1, \xi_2) \widehat{u}(\xi_1) \widehat{u}(\xi_2) e^{ix \cdot (\xi_1 + \xi_2)}$$

The classical Birkhoff normal form procedure consists of trying to eliminate the quadratic terms in eq. (66). One then looks for a change of variables

$$u = w + \mathcal{F}(w) \quad \text{where} \\ \mathcal{F}(w) := \sum_{\xi_1, \xi_2} f(\xi_1, \xi_2) \widehat{w}(\xi_1) \widehat{w}(\xi_2) e^{ix \cdot (\xi_1 + \xi_2)}. \quad (67)$$

A direct calculation shows that the equation of the variable  $w$  takes the form

$$\begin{aligned} \partial_t w + i\Lambda(D)w + \sum_{\xi_1, \xi_2} (i(\Lambda(\xi_1 + \xi_2) - \Lambda(\xi_1) - \Lambda(\xi_2))f(\xi_1, \xi_2) + q(\xi_1, \xi_2)) \widehat{w}(\xi_1) \widehat{w}(\xi_2) e^{ix \cdot (\xi_1 + \xi_2)} \\ + \text{cubic terms} = 0 \end{aligned}$$

So if  $\Lambda(\xi_1 + \xi_2) - \Lambda(\xi_1) - \Lambda(\xi_2) \neq 0$  for any  $\xi_1, \xi_2$ , one can define (at least formally)

$$f(\xi_1, \xi_2) := - \frac{q(\xi_1, \xi_2)}{i(\Lambda(\xi_1 + \xi_2) - \Lambda(\xi_1) - \Lambda(\xi_2))} \quad (68)$$

in such a way to cancel out the quadratic terms from the equation. This procedure can be formally iterated, provided one is able to prove suitable lower bounds for the *divisors*

$$\Lambda(\xi_1 + \dots + \xi_N) - \Lambda(\xi_1) - \dots - \Lambda(\xi_N)$$

when this divisor is different from zero and replacing bilinear forms by multilinear forms one defines

$$f(\xi_1, \dots, \xi_N) := - \frac{q(\xi_1, \dots, \xi_N)}{i(\Lambda(\xi_1 + \dots + \xi_N) - \Lambda(\xi_1) - \dots - \Lambda(\xi_N))} \quad (69)$$

In the framework of the Water Waves equation, there are two problems in defining rigorously the normal form transformations arising in the Birkhoff normal form procedure: (1) the quadratic form  $Q$  in eq. (66), is unbounded implying that the coefficients  $q(\xi_1, \xi_2)$  diverge with some power of  $|\xi_1|, |\xi_2|$  of the same order as the diagonal part  $\Lambda(D)$  (2) The divisors eq. (69) could accumulate to zero leading to further *loss of derivatives* in the estimate of  $f(\xi_1, \dots, \xi_N)$ . This problem has been fully overcome (at least for 2d periodic Water Waves) in Berti and Delort (2018) at any order, meaning that for a large set of surface tension parameter  $\kappa > 0$ , for initial data of size  $\varepsilon$  small enough, the solution stays smooth and small of order  $O(\varepsilon)$  over a time interval of length  $O(\varepsilon^{-N})$ . The fundamental technical ingredient is para-differential calculus, which is a generalization of pseudo-differential calculus to symbols with limited spatial smoothness. We refer to the book Metivier (2008) for an exhaustive presentation of this theory. Without any restriction on the parameters, a finite number of quadratic resonances could appear. Such resonances are called Wilton ripples. This problem has been studied in Berti et al. (2020a), in which the authors proved a lifespan of order  $O(\varepsilon^{-2})$  for the solutions of gravity capillary Water Waves for any value of the surface tension  $\kappa > 0$ . For the pure gravity Water Waves equations, a quadratic Birkhoff normal form step has been performed in Wu (2009), Ionescu and Pusateri (2015), Alazard and Delort (2015a), Hunter et al. (2016), and Craig and Sulem (2016). The integrability of 2d pure gravity Water Waves up to order four has been formally proved in Dyachenko and Zakharov (1994). They basically show that there exists a (formal) change of coordinates which

reduces the Water Waves equations to another one which is integrable up to monomials of order four. It was a longstanding open problem to prove whether the integrability up to order four of the pure gravity Water Waves could be made rigorous. This Dyachenko-Zakharov conjecture has been recently solved in Berti et al. (2018).

In the next section “[Birkhoff Normal Form and Longtime Existence for Gravity Capillary Water Waves](#),” we shall describe the Birkhoff normal form theory for 2d gravity capillary Water Waves developed in Berti and Delort (2018). Concerning the 2d pure gravity Water Waves, we shall describe in section “[The Dyachenko-Zakharov Conjecture for Pure Gravity Water Waves](#)” how the Dyachenko-Zakharov conjecture has been solved in Berti et al. (2018).

### Birkhoff Normal Form and Longtime Existence for Gravity Capillary Water Waves

In this section we present rigorously the result obtained in Berti and Delort (2018), concerning 2d gravity capillary Water Waves, see eq. 21. The *almost global existence result* is the following

**Theorem 5.1 (Berti and Delort 2018)** *There is a zero measure set  $\mathcal{N} \subseteq (0, +\infty)$  such that for any value of the surface tension  $\kappa \in (0, +\infty) \setminus \mathcal{N}$ , for any  $N \in \mathbb{N}$ , there is  $s_0 \equiv s_0(N) > 0$  such that for any  $s \geq s_0$ , for any even initial data  $(\eta_0, \psi_0) \in H^{s+\frac{1}{4}}(\mathbb{T}) \times H^{s-\frac{1}{4}}(\mathbb{T})$ , with*

$$\|\eta_0\|_{s+\frac{1}{4}} + \|\psi_0\|_{s-\frac{1}{4}} \leq \varepsilon$$

*with  $\varepsilon \equiv \varepsilon(s) \ll 1$  small enough, there exists a unique solution  $(\eta, \psi) \in C^0([-T_\varepsilon, T_\varepsilon], H^{s+\frac{1}{4}}(\mathbb{T}) \times H^{s-\frac{1}{4}}(\mathbb{T}))$  of the gravity capillary Water Waves eq. (21), with*

$$T_\varepsilon = O(\varepsilon^{-N}) \quad \text{and} \quad \sup_{t \in [-T_\varepsilon, T_\varepsilon]} \left( \|\eta(t, \cdot)\|_{s+\frac{1}{4}} + \|\psi(t, \cdot)\|_{s-\frac{1}{4}} \right) \lesssim \varepsilon.$$

As we already mentioned, the proof strongly relies on paradifferential calculus. The first step of

the proof is actually to construct a *modified energy* for the eq. (21),

$$(\eta, \psi) := \Phi(\eta, \psi)[u, \bar{u}]$$

where  $\Phi(\eta, \psi)$  is a bounded, linear and invertible transformation (depending nonlinearly on  $(\eta, \psi)$ )  $H^s(\mathbb{T}) \times H^s(\mathbb{T}) \rightarrow H^{s+\frac{1}{4}}(\mathbb{T}) \times H^{s-\frac{1}{4}}(\mathbb{T})$  for  $s \gg 0$  large enough, such that the equation in the new complex coordinate  $u$  takes the form

$$\partial_t u + iT(D)u + i\lambda(u; t, D)[u] + \mathcal{R}(u) = 0 \quad (70)$$

where  $T(D) := \sqrt{|D|(1 - \kappa \partial_{xx})}$ ,  $\lambda(u, t, D)$  is a real Fourier multiplier depending nonlinearly on  $u$ , namely

$$\lambda(u, t, D)[u] = \sum_{j \in \mathbb{Z}} \lambda(u; t, \xi) \widehat{u}(\xi) e^{i \xi x}, \quad \lambda(u; t, \xi) \in \mathbb{R} \quad (71)$$

and the remainder is smoothing, namely

$$H^s(\mathbb{T}) \rightarrow H^{s+\sigma}(\mathbb{T}), \quad u \mapsto \mathcal{R}(u) \quad (72)$$

with  $\sigma \gg 0$  large enough. This step requires to construct normal form transformations by using paradifferential calculus and, roughly speaking, it is a nonlinear analogue of the procedure developed in the KAM result obtained in Berti and Montalto (2020) and described in section “Some Ideas of the Proof.” Then, one can implement a Birkhoff normal form procedure on the eq. (70), in order to normalize the smoothing remainder  $\mathcal{R}(u)$ . For most values of the surface tension  $\kappa \in (0, +\infty)$ , one can verify for any  $N \in \mathbb{N}$ , non-resonance conditions

$$\left| \sum_{i=1}^N \sigma_{j_i} w_{j_i}(\kappa) \right| \geq \frac{\gamma}{\left( \sum_{i=1}^N j_i \right)^\tau} \text{ if } N \text{ is odd}$$

$$\left| \sum_{i=1}^N \sigma_{j_i} w_{j_i}(\kappa) \right| \geq \frac{\gamma}{\left( \sum_{i=1}^N j_i \right)^\tau} \text{ if } N \text{ is even} \quad (73)$$

when  $\{j_i, \dots, j_N\} \neq \{k_1, k_1, k_2, k_2, \dots, k_{N/2}, k_{N/2}\}$ ,  $\{\sigma_{j_i}, \dots, \sigma_{j_N}\} \neq \{\sigma_{k_1}, -\sigma_{k_1}, \dots, \sigma_{N/2}, -\sigma_{N/2}\}$  for some  $\tau = \tau(N) \gg 0$  large enough, where  $\sigma_{j_i} \in \{+, -\}$  and  $\omega_j(\kappa) = \sqrt{j(1 + \kappa j^2)}$ ,  $j \geq 1$ .

Then, by choosing  $\sigma \gg \tau$  in eq. (72), the loss of the small divisors in eq. (73), is compensated by the smoothing effect of the remainder eq. (72). Hence, one can construct a Birkhoff transformation (as it is illustrated in eqs. (67)–(69)) which is bounded and invertible on Sobolev spaces  $H^s(\mathbb{T})$ ,  $s \gg 0$ . After  $N$  steps of Birkhoff normal form, one gets an equation of the form

$$\partial_t u + iT(D)u + iz(u; t, D)[u] + Q_N(u) = 0 \quad (74)$$

where  $z$  has the same form as  $\lambda$  in eq. (71) and the remainder  $H^s(\mathbb{T}) \rightarrow H^s(\mathbb{T})$ ,  $u \mapsto Q_N(u)$  is bounded and satisfies

$$\|Q_N(u)\|_s \lesssim \|u\|_s^{N+1}.$$

Due to the fact that  $z(u; t, D)$  is a real Fourier multiplier, the *normal form dynamics* (i.e., eq. (74) by neglecting  $Q_N(u)$ )

$$\partial_t u + iT(D)u + iz(u; t, D)[u] = 0$$

satisfies

$$\|u(t, \cdot)\|_s = \|u(0, \cdot)\|_s, \forall t.$$

It is immediate then to verify by energy estimates that any solution of eq. (74), satisfies for any  $t \in [-T, T]$ ,

$$\|u(t, \cdot)\|_s \leq \|u(0, \cdot)\|_s + C(s) \int_0^T \|u(\tau, \cdot)\|_s^{N+1} d\tau$$

which implies that if  $\|u_0\|_s \leq \varepsilon$ , then  $\|u(t, \cdot)\|_s \lesssim \varepsilon$  for any  $t \in [-T, T]$  with  $T = O(\varepsilon^{-N})$ .

### The Dyachenko-Zakharov Conjecture for Pure Gravity Water Waves

In this section, we consider the pure gravity Water Waves in infinite depth, namely the system

$$\begin{cases} \partial_t \eta = G(\eta) \psi, \\ \partial_t \psi + \eta + \frac{1}{2} \psi_x^2 - \frac{1}{2} \frac{(G(\eta) \psi + \eta_x \psi_x)^2}{1 + \eta_x^2} = 0 \end{cases} \quad (75)$$

with Hamiltonian

$$H(\eta, \psi) := \frac{1}{2} \int_{\mathbb{T}} G(\eta) |\psi|^2 dx + \frac{1}{2} \int_{\mathbb{T}} \eta^2 dx. \quad (76)$$

In infinite depth, the Dirichlet Neumann operator at flat surface  $\eta = 0$  is given by  $G(0) = |D|$  and by defining the complex scalar unknown

$$u := \frac{1}{\sqrt{2}} |D|^{-\frac{1}{4}} \eta + \frac{i}{\sqrt{2}} |D|^{\frac{1}{4}} \psi, \quad (77)$$

the eq. (75) takes the form

$$\partial_t u + i |D|^{1/2} u = Q(u)$$

where  $Q(u)$  is a fully nonlinear, quadratic vector field which contains up to first order derivatives of  $u$ . The dispersion relation is then given by

$$w(j) = \sqrt{j}, j \geq 1. \quad (78)$$

In Dyachenko and Zakharov (1994) (see also Craig and Worfolk 1995)) Dyachenko and Zakharov proved, at least at a formal level, the integrability of eq. (75) up to order four, namely they constructed a formal Birkhoff transformation which transform the system eq. (75) into an equation which is integrable up to quartic terms  $O(|\eta, \psi|^4)$ . Since their construction is only formal no information is given on the stability of small solutions of order  $O(\varepsilon)$  over a time interval of order  $O(\varepsilon^{-3})$ . At a formal level, the Birkhoff normal form procedure would be possible in the absence of nontrivial 3-waves and 4-waves resonances, namely integer solutions of

$$\begin{aligned} \sigma_1 \omega(j_1) + \sigma_2 \omega(j_2) + \sigma_3 \omega(j_3) &= 0, \\ \sigma_1 j_1 + \sigma_2 j_2 + \sigma_3 j_3 &= 0, \sigma_i \in \{+, -\}, j_i \in \mathbb{N}_+ \end{aligned} \quad (79)$$

and

$$\begin{aligned} \sigma_1 \omega(j_1) + \sigma_2 \omega(j_2) + \sigma_3 \omega(j_3) + \sigma_4 \omega(j_4) &= 0, \\ \sigma_1 j_1 + \sigma_2 j_2 + \sigma_3 j_3 + \sigma_4 j_4 &= 0, \sigma_i \in \{+, -\}, j_i \in \mathbb{N}_+ \end{aligned} \quad (80)$$

which do not appear in pairs with corresponding opposite signs. It turns out that for the dispersion relation eq. (78), there are no three wave interactions eq. (79). On the other hand, the property eq. (80) is not satisfied by the dispersion relation eq. (78). Indeed, as shown in Dyachenko and Zakharov (1994), there are many solutions of eq. (80). For example, if  $\sigma_1 = \sigma_3 = 1 = -\sigma_2 = -\sigma_4$ , in addition to the trivial solutions  $(n_1, n_2, n_3, n_4) = (k, k, j, j)$ , there is the two parameter family of solutions, called *Benjamin-Feir resonances*,

$$\begin{aligned} \bigcup_{\lambda \in \mathbb{Z} \setminus \{0\}, b \in \mathbb{N}} \{ &n_1 = -\lambda b^2, n_2 = \lambda(b+1)^2, \\ &n_3 = \lambda(b^2 + b + 1)^2, n_4 = \lambda(b+1)^2 b^2 \}. \end{aligned} \quad (81)$$

Applying a purely formal reduction to Birkhoff normal form up to order four, the trivial resonances give rise to benign integrable monomials of the form  $|z_k|^2 |z_j|^2$ , whereas the Benjamin-Feir resonances could give nonintegrable monomials of the form  $z_{-lb^2} \overline{z_{l(b+1)^2}} z_{l(b^2+b+1)} \overline{z_{l(b+1)^2 b^2}} + \text{c.c.}$

A striking property proved in Dyachenko and Zakharov (1994), see also Craig and Worfolk (1995), Craig and Sulem (2016), is that the coefficients of the formal Birkhoff Hamiltonian which are supported on eq. (81), are actually zero. In particular, one has the following

**Theorem 5.2 (Formal Integrability at Order Four)** *There exists a formal transformation  $\Phi$  such that the truncation of  $H \circ \Phi$  at order four of homogeneity is given by the Hamiltonian  $H_{\text{ZD}}$  which has the form*

$$H_{\text{ZD}} = H_{\text{ZD}}^{(2)} + H_{\text{ZD}}^{(4)}, \quad H_{\text{ZD}}^{(2)}(z, \bar{z}) := \frac{1}{2} \int_{\mathbb{T}} |D|^{\frac{1}{4}} |z|^2 dx, \quad (82)$$



with

$$\begin{aligned}
 H_{\text{ZD}}^{(4)}(z, \bar{z}) := & \frac{1}{4\pi} \sum_{k \in \mathbb{Z}} |k|^3 \left( |z_k|^4 - 2|z_k|^2 |z - k|^2 \right) \\
 & + \frac{1}{\pi} \sum_{\substack{k_1, k_2 \in \mathbb{Z}, \text{sign}(k_1) = \text{sign}(k_2) \\ |k_2| < |k_1|}} \\
 & |k_1| |k_2|^2 \left( -|z_{-k_1}|^2 |z_{k_2}|^2 + |z_{k_1}|^2 |z_{k_2}|^2 \right)
 \end{aligned} \quad (83)$$

where  $z_k$  denotes the  $k$ -th Fourier coefficient of the function  $z$ . Moreover,  $H_{\text{ZD}}$  is integrable, since it depends only on the action variables  $I_k := |z_k|^2$  which are constants of the motion. In particular, its flow preserves all Sobolev norms.

This result is a purely formal calculation, and no actual relation is established between the flow of  $H$  (which is well-posed for short times) and that of  $H \circ \Phi$  or  $H_{\text{ZD}}$ . This is the goal of the result proved in Berti et al. (2018).

**Theorem 5.3 (The Dyachenko-Zakharov Conjecture: Cubic Lifespan for Pure Gravity Water Waves in Infinite Depth)** *There exists  $s_0 > 0$  such that, for all  $s \geq s_0$ , there is  $\varepsilon > 0$  small enough such that, for any initial data  $(\eta_0, \psi_0)$  satisfying*

$$\|\eta_0\|_s + \|\psi_0\|_s \leq \varepsilon, \quad \int_{\mathbb{T}} \eta_0(x) dx = 0, \quad (84)$$

*there exists a unique classical solution  $(\eta, \psi) \in C^0([-T_\varepsilon, T_\varepsilon], H^s(\mathbb{T}) \times H^s(\mathbb{T}))$  of the Water Waves system eq. (75) with initial condition  $(\eta, \psi)|_{t=0} = (\eta_0, \psi_0)$  with*

$$T_\varepsilon = O(\varepsilon^{-3}), \quad (85)$$

satisfying

$$\sup_{[-T_\varepsilon, T_\varepsilon]} (\|\eta(t, \cdot)\|_s + \|\psi(t, \cdot)\|_s) \leq \varepsilon, \quad \int_{\mathbb{T}} \eta(t, x) dx = 0. \quad (86)$$

The latter theorem is proved by using the same technology developed in Berti and Delort (2018) and explained in section “Birkhoff Normal Form and Longtime Existence for Gravity Capillary Water Waves.” By combining paradifferential calculus and Birkhoff normal form techniques, the authors transform by means of a bounded and invertible map on  $H^s$ ,  $s \gg 0$ , the Water Waves equation into another equation having the form

$$\partial_t z + i|D|^{\frac{1}{2}} z + \mathcal{X}^{\text{res}}(z) + \mathcal{X}_{\geq 4}(z) = 0 \quad (87)$$

with the following properties.  $\mathcal{X}_{\geq 4}(z)$  is a quartic terms and satisfies the energy estimate

$$\text{Re} \left( \int_{\mathbb{T}} |D|^s \mathcal{X}_{\geq 4}(z) \cdot \overline{|D|^s z} dx \right) \lesssim \|z\|_s^5.$$

The vector field  $\mathcal{X}^{\text{res}}$  is cubic and it is supported only by the resonant monomials (see eq. (80)). More precisely it has the form

$$\begin{aligned}
 \mathcal{X}^{\text{res}}(z) = & \sum_{\substack{n = \sigma_1 n_1 + \sigma_2 n_2 + \sigma_3 n_3 \\ \omega(n) = \sigma_1 \omega(n_1) + \sigma_2 \omega(n_2) + \sigma_3 \omega(n_3) \\ (\mathcal{X}_{\sigma_1, \sigma_2, \sigma_3}(n_1, n_2, n_3) z_{n_1}^{\sigma_1} z_{n_2}^{\sigma_2} z_{n_3}^{\sigma_3}) e^{inx} \\ \mathcal{X}_{\sigma_1, \sigma_2, \sigma_3}(n_1, n_2, n_3) \in \mathbb{C}.
 \end{aligned}$$

In the latter formula,  $\sigma_1, \sigma_2, \sigma_3 \in \{+, -\}$  and we use the notation  $z_n^+ := z_n$  and  $z_n^- := \overline{z_n}$ .

Then, the last main step is an a posteriori *identification argument* to prove that the cubic resonant terms  $\mathcal{X}^{\text{res}}(z)$  in eq. (87) are uniquely determined and coincide with the Hamiltonian vector field generated by the fourth order Birkhoff normal form Hamiltonian  $H_{\text{ZD}}^{(4)}$  in eq. (83), that is,

$$\mathcal{X}^{\text{res}}(z) = i\partial_{\bar{z}} H_{\text{ZD}}^{(4)}.$$

Note that the identification of the linear terms in  $z$  is obvious by an easy calculation  $i|D|^{\frac{1}{2}} z = i\partial_{\bar{z}} H_{\text{ZD}}^{(2)}$  (see eq. (82)). We also mention that the uniqueness of the normal form is based on the absence of cubic resonances. This argument then shows that the normal form vector field  $i|D|^{\frac{1}{2}} z +$

$\chi^{\text{res}}(z)$  obtained in eq. (87) is integrable. Theorem 5.3 is then deduced by an energy estimate for the eq. (87). The rigorous integrability of the pure gravity Water Waves equation up to order four is somehow optimal. Indeed, in Dyachenko et al. (1995) the authors proved the nonintegrability of the Birkhoff normal form of order five.

## Future Developments

In this last section we shall describe some open problems on the theory of Water Waves. From what we said in sections “KAM Results for Water Waves” and “Longtime Existence for Periodic Water Waves,” it turns out that the perturbative approach to 2d Water Waves is nowadays quite complete. So the next further challenge is to understand how perturbation theory works on 3d Water Waves.

### KAM theory for 3d periodic Water Waves.

A breakthrough problem would be to develop the KAM results for 3d Water Waves, in both cases gravity and gravity-capillary Water Waves. Existence of small amplitude 3d traveling gravity-capillary water waves solutions with space periodic boundary conditions has been proved in Craig and Nicholls (2000) (it is not a small divisor problem) and in Iooss and Plotnikov (2009)–Iooss and Plotnikov (2011) in the case of zero surface tension (in such a case it is a small divisor problem). They found traveling waves oscillating with two frequencies  $(\omega_1, \omega_2)$  of the form  $u(x_1 - \omega_1 t, x_2 - \omega_2 t)$  (where  $u(v_1, v_2)$  is periodic in both variables). These traveling wave solutions are *stationary in a moving frame*. This fact simplifies the analysis of the linearized equations and the small divisor’s problem. It would be a very challenging problem to prove the existence of quasi-periodic solutions with an arbitrary number of frequencies. It is worth to mention that the extension of the KAM to PDEs in higher space dimension with unbounded perturbations is a very difficult matter, even on simpler models. Only few results are available on physically relevant models (see for instance Baldi and Montalto (2021), for the 3D Euler equation with a time quasi-periodic external force).

**Long time existence for 3d periodic Water Waves.** Concerning the extension to 3d periodic Water Waves of the normal form results illustrated in section “Longtime Existence for Periodic Water Waves,” the only result available is Ionescu and Pusateri (2019), in which the authors proved that for small data of size  $O(\varepsilon)$ , the solutions of gravity capillary Water Waves stay regular and small  $O(\varepsilon)$  for a time interval of size  $O\left(\varepsilon^{-\left(\frac{5}{3}+\eta\right)}\right)$  for some  $\eta \ll 1$  small enough. This time of existence is strictly better than the lifespan provided by the local existence. It would be very challenging then to investigate the possibility of almost global existence results for 3d periodic Water Waves. The major difficulties are: (1) the resonances in higher space dimension could be much more complicated than in 1d. (2) The extension in higher space dimension of the normal form techniques based on microlocal analysis.

It would be also interesting to investigate some instability phenomena. For instance, to construct *turbulent solutions*, namely solutions whose Sobolev norm becomes very large after a certain time  $T$ , in the spirit of Colliander et al. (2010).

## Bibliography

- Alazard T, Baldi P (2015) Gravity capillary standing Water Waves. Arch Rat Mech Anal 217(3):741–830
- Alazard T, Delort J-M (2015a) Sobolev estimates for two dimensional gravity water waves. Astérisque 374: viii + 241
- Alazard T, Delort J-M (2015b) Global solutions and asymptotic behavior for two dimensional gravity Water Waves. Ann Sci Éc Norm Supér 48(5): 1149–1238
- Alazard T, Burq N, Zuily C (2011) On the water-wave equations with surface tension. Duke Math J 158: 413–499
- Alazard T, Burq N, Zuily C (2014) On the Cauchy problem for gravity water waves. Invent Math 198:71–163
- Baldi P, Montalto R (2021) Quasi-periodic incompressible Euler flows in 3D. Adv Math 384:107730. <https://doi.org/10.1016/j.aim.2021.107730>
- Baldi P, Berti M, Montalto R (2014) KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation. Math Ann 359(1–2):471–536
- Baldi P, Berti M, Montalto R (2016) KAM for autonomous quasi-linear perturbations of KdV. Ann Inst H Poincaré Analyse Non Lin 33(6):1589–1638

- Baldi P, Berti M, Haus E, Montalto R (2018) Time quasi-periodic gravity Water Waves in finite depth. *Invent Math* 214:739911
- Bambusi D, Berti M, Magistrelli E, Degenerate KAM (2011) Theory for partial differential equations. *J Diff Equ* 250(8):3379–3397
- Berti M (2016) KAM for PDEs. *Boll Unione Mat Ital* 9: 115–142
- Berti M, Delort J-M (2018) Almost global solutions of capillary-gravity Water Waves equations on the circle (UMI Lecture Notes 2018). ISBN 978-3-319-99486-4
- Berti M, Montalto R (2020) Quasi-periodic standing wave solutions for gravity capillary Water Waves. *Mem Am Math Soc* 263:1273
- Berti M, Feola R, Pusateri F (2018) Birkhoff normal form and long time existence for periodic gravity Water Waves. *Comm Pure Appl Math* arXiv:1810.11549
- Berti M, Feola R, Franzoi L (2020a) Quadratic life span of periodic gravity-capillary Water Waves. *Water Waves*. <https://doi.org/10.1007/s42286-020-00036-8>
- Berti M, Franzoi L, Maspero A (2020b) Traveling quasi-periodic Water Waves with constant vorticity. Preprint arXiv:2004.08905
- Bourgain J (1994) Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE. *Int Math Res Not* 11:475
- Christodoulou D, Lindblad H (2000) On the motion of the free surface of a liquid. *Comm Pure Appl Math* 53(12): 1536–1602
- Colliander J, Keel M, Staffilani G, Takaoka H, Tao T (2010) Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. *Invent Math* 181:39–113
- Constantin A (2011) Nonlinear Water Waves with applications to wave-current interaction and Tsunamis. CBMS-NSF Regional Conference Series in Applied Mathematics, vol 81. SIAM, Philadelphia
- Constantin A, Strauss W (2004) Exact steady periodic water waves with vorticity. *Comm Pure Appl Math* 57(4):481–527
- Craig W (1985) An existence theory for Water Waves and the Boussinesq and Kortewegde Vries scaling limits. *Commun Partial Differ Equ* 10(8):787–1003
- Craig W, Nicholls D (2000) Traveling two and three dimensional capillary gravity Water Waves. *SIAM J Math Anal* 32(2):323–359
- Craig W, Sulem C (1993) Numerical simulation of gravity waves. *J Comput Phys* 108(1):73–83
- Craig W, Sulem C (2016) Mapping properties of normal forms transformations for Water Waves. *Boll Unione Mat Italiana* 9(2):289–318
- Craig W, Wayne CE (1993) Newton's method and periodic solutions of nonlinear wave equation. *Comm Pure Appl Math* 46:1409–1498
- Craig W, Worfolk P (1995) An integrable normal form for Water Waves in infinite depth. *Phy D* 84(3–4):513–531
- Deng Y, Ionescu AD, Pausader B, Pusateri F (2017) Global solutions for the 3D gravity-capillary Water Waves system. *Acta Math* 219(2):213–402
- Dubreil-Jacotin M-L (1934) Sur la dtermination rigoureuse des ondes permanentes priodiques dampleur finie. *J Math Pures Appl* 13:217–291
- Dyachenko AI, Zakharov VE (1994) Is free-surface hydrodynamics an integrable system? *Phys Lett A* 190: 144–148
- Dyachenko AI, Lvov YV, Zakharov VE (1995) Five-wave interaction on the surface of deep fluid. *Phys D* 87: 233–261
- Feola R, Giuliani F (2020) Quasi-periodic traveling waves on an infinitely deep fluid under gravity. Preprint arXiv:2005.08280
- Gaeta G (2014) Nonlinear dynamics, symmetry and perturbation theory. In: *Encyclopedia of complexity and systems science*. Springer, New York. <https://doi.org/10.1007/978-3-642-27737-5>
- Germain P, Masmoudi N, Shatah J (2012) Global solutions for the gravity water waves equation in dimension 3. *Ann Math* 175(2):691–754
- Gerstner F (1802) Theorie der Wellen. *Abh Knigl Böhm Ges Wiss*
- Goyon R (1958) Contribution la thorie des houles. *Ann Sci Univ Toulouse* 22:1–55
- Hunter J, Ifrim M, Tataru D (2016) Two dimensional Water Waves in holomorphic coordinates. *Comm Math Phys* 346:483–552
- Ifrim M, Tataru D (2016) Two dimensional Water Waves in holomorphic coordinates II: global solutions. *Bull Soc Math France* 144:369–394
- Ionescu A, Pusateri F (2015) Global solutions for the gravity Water Waves system in 2d. *Invent Math* 199(3):653–804
- Ionescu A, Pusateri F (2016) Global analysis of a model for capillary Water Waves in 2D. *Comm Pure Appl Math* 69(11):2015–2071
- Ionescu A, Pusateri F (2018a) Global regularity for 2d Water Waves with surface tension. *Mem Am Math Soc* 256(1227):1–136
- Ionescu A, Pusateri F (2018b) Recent advances on the global regularity for Water Waves. *Philos Trans Roy Soc A* 376(2111):20170089
- Ionescu A, Pusateri F (2019) Long-time existence for multi-dimensional periodic Water Waves. *Geom Funct Anal* 29(3):811–870
- Iooss G, Plotnikov P (2005) Existence of multimodal standing gravity waves. *J Math Fluid Mech* 7:349–364
- Iooss G, Plotnikov P (2009) Small divisor problem in the theory of three-dimensional water gravity waves. *Mem Am Math Soc* 200(940):viii+128
- Iooss G, Plotnikov P (2011) Asymmetrical tridimensional traveling gravity waves. *Arch Rat Mech Anal* 200(3): 789–880
- Iooss G, Plotnikov P, Toland J (2005) Standing waves on an infinitely deep perfect fluid under gravity. *Arch Rat Mech Anal* 177(3):367–478

- Keady G, Norbury J (1978) On the existence theory for irrotational water waves. *Math Proc Camb Philos Soc* 83(1):137–157
- Kuksin S (1987) Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum. *Funktsional Anal Prilozhen* 21(3):22–37, 95
- Lannes D (2005) Well-posedness of the Water Waves equations. *J Am Math Soc* 18(3):605–654
- Lannes D (2013) The Water Waves problem: mathematical analysis and asymptotics. *Mathematical Surveys and Monographs*, vol 188. American Mathematical Society, Providence
- Levi-Civita T (1925) D'etermination rigoureuse des ondes permanentes d' ampleur finie. *Math Ann* 93:264–314
- Martin C (2013) Local bifurcation and regularity for steady periodic capillary-gravity water waves with constant vorticity. *Nonlinear Anal Real World Appl* 14(1):131–149
- McLeod JB (1997) The Stokes and Krasovskii conjectures for the wave of greatest height. *Stud Appl Math* 98(4): 311–333
- Métivier G (2008) Para-differential Calculus and applications to the Cauchy problem for nonlinear systems. *Pubblicazioni Scuola Normale Pisa*, Pisa, p 5
- Nalimov VI (1974) The Cauchy-Poisson problem. *Dinamika Splosn. Sredy Vyp. 18 Dinamika Zidkost. so Svobod. Granicami* 10–210:254
- Nekrasov AI (1921) On steady waves. *Izv. Ivanovo-Voznesensk. Politekhn* 3:52–65
- Nirenberg L (2001) Topics in nonlinear functional analysis. *Courant lecture notes*, vol 6. American Mathematical Society, Providence
- Plotnikov P, Toland J (2001) Nash-Moser theory for standing Water Waves. *Arch Rat Mech Anal* 159(1):1–83
- Pöschel J (1996) A KAM-theorem for some nonlinear partial differential equations. *Ann Scuola Norm Sup Pisa Cl Sci* 23(1):119–148
- Saranen J, Vainikko G (2002) Periodic integral and Pseudodifferential equations with numerical approximation. *Springer monographs in mathematics*. Springer, New York
- Shatah J, Zeng C (2011) Local well-posedness for the fluid interface problem. *Arch Ration Mech Anal* 199(2): 653–705
- Stoker JJ (1992) Water waves. The mathematical theory with applications. *Wiley Classics Library Edition* Published, New York
- Stokes G (1847) On the theory of oscillatory waves. *Trans Cambridge Phil Soc* 8:441–455
- Struik D (1926) Dtermination rigoureuse des ondes irrotationnelles p'eriodes dans un canal profondeur finie. *Math Ann* 95:595–634
- Taylor ME (1991) Pseudodifferential operators and nonlinear PDEs. *Progress in mathematics*. Birkhäuser, Boston
- Toland J (1978) On the existence of a wave of greatest height and Stokes conjecture. *Proc Roy Soc London A* 363(1715):469–485
- Wahlén E (2006) Steady periodic capillary-gravity waves with vorticity. *SIAM J Math Anal* 38:921–943
- Wang X (2018) Global infinite energy solutions for the 2D gravity Water Waves system. *Comm Pure Appl Math* 71(1):90–162
- Wayne E (1990) Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory. *Comm Math Phys* 127:479–528
- Whitham GB (1999) Linear and nonlinear waves. *Wiley*. <https://doi.org/10.1002/9781118032954>
- Wu S (1997) Well-posedness in Sobolev spaces of the full Water Waves problem in 2-D. *Invent Math* 1(130): 39–72
- Wu S (2009) Almost global wellposedness of the 2-D full water wave problem. *Invent Math* 1(177):45–135
- Wu S (2011) Global well-posedness of the 3-D full water wave problem. *Invent Math* 1(184):125–220
- Yosihara H (1982) Gravity waves on the free surface of an incompressible perfect fluid of finite depth. *Publ Res Inst Math Sci* 18:49–96
- Zakharov V (1968) Stability of periodic waves of finite amplitude on the surface of a deep fluid. *J Appl Mech Tech Phys* 9(2):190–194
- Zeidler E (1971) Existenzbeweis fr cnoidal waves unter Bercksichtigung der Oberflchen spannung. *Arch Rational Mech Anal* 41:81–107
- Zeidler E (1973) Existenzbeweis fr permanente Kapillar-Schwerewellen mit allgemeinen Wirbelverteilungen. *Arch Ration Mech Anal* 50:3472

# Periodic Rogue Waves and Perturbation Theory

F. Coppini<sup>1,2</sup>, P. G. Grinevich<sup>3,4,5</sup> and P. M. Santini<sup>1</sup>

<sup>1</sup>Dipartimento di Fisica, Università di Roma, “La Sapienza”, Rome, Italy

<sup>2</sup>Department of Mathematics, Physics and Electrical Engineering, Northumbria University Newcastle, Newcastle upon Tyne, UK

<sup>3</sup>Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia

<sup>4</sup>L.D. Landau Institute for Theoretical Physics, Chernogolovka, Moscow region, Russia

<sup>5</sup>Department of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia

## Article Outline

Glossary

Definition of the Subject

Introduction

The Finite Gap Method and the Periodic NLS

Cauchy Problem of the Rogue Waves, for a Finite Number of Unstable Modes

The Solution of the Cauchy Problem in Terms of Elementary Functions

RW Perturbation Theory

Future Directions

References

## Glossary

**Divisor** The points of the Dirichlet spectrum on the Riemann surface for the Bloch eigenfunctions, playing the role of the angle variables in the finite-gap method.

**Finite-gap solutions** Periodic (quasiperiodic) solutions of soliton equations with the

following property: the Bloch eigenfunctions for the auxiliary linear operator are meromorphic on a Riemann surface with a finite number of branch and double points. They may be treated as nonlinear analogs of finite Fourier series solutions.

**Modulation instability** a mechanism through which a perturbation of a periodic waveform grows due to nonlinearity of the media. It is considered as the main mechanism for rogue wave generation.

**Rogue (anomalous, freak) waves** extreme waves of anomalously large amplitude with respect to the surrounding waves, arising apparently from nowhere and disappearing without leaving any trace.

**Soliton equations** nonlinear partial differential equations arising as the compatibility condition for a pair of auxiliary linear operators, and integrable (linearizable) using the spectral transform for these operators. Soliton equations usually possess infinitely many integrals of motion in involution and wide classes of exact solutions.

## Definition of the Subject

Rogue waves (RWs) are extreme waves of anomalously large amplitude with respect to the surrounding waves, arising apparently from nowhere and disappearing without leaving any trace. The simplest basic model describing the generation of RWs, due to modulation instability in nonlinear media, is the self-focusing Nonlinear Schrödinger (NLS) equation. Here we study the problem in the spatially periodic setting, using the finite gap method, for initial perturbations of the unstable background solution, in the case of a finite number of unstable modes. In addition, we construct a perturbation theory allowing one to study the effects of a small perturbation of the NLS equation on the dynamics of periodic rogue waves.



## Introduction

Rogue waves (RWs), also called anomalous, freak waves, are extreme waves of anomalously large amplitude with respect to the surrounding waves, arising apparently from nowhere and disappearing without leaving any trace. Deep seawater is the first environment where RWs have been studied, and the term RW was first coined by oceanographers (an RW can exceed the height of 20 m and can be very dangerous). Although the existence of RWs was notorious even to the ancients, and it is a recurring theme in the history of literature, the first scientific observation and measurement of an RW has been made only in 1995 at the Draupner oil platform in the North Sea (Haver 1995). RWs are not confined to oceanography, and their presence is ubiquitous in nature: they have been observed or predicted in nonlinear optics (Solli et al. 2007; Kibler et al. 2010, 2012; Akhmediev et al. 2013; Liu et al. 2015), in Bose-Einstein condensates (Bludov et al. 2009), in plasma physics (Bailung et al. 2011; Moslem et al. 2011), in acoustic turbulence in superfluid Helium (Ganshin et al. 2008), and in other areas of physics.

Modulation instability (MI), first observed in nonlinear optics (Bespalov and Talanov 1966) and ocean waves (Benjamin and Feir 1967; Zakharov 1968), is considered as the main physical mechanism for the appearance of RWs in nature (Henderson et al. 1999; Dysthe and Trulsen 1999; Osborne et al. 2000; Kharif and Pelinovsky 2011; Kharif et al. 2009; Onorato et al. 2013; Zakharov and Ostrovsky 2009), and the self-focusing Nonlinear Schrödinger (NLS) equation

$$iu_t + u_{xx} + 2|u|^2u = 0, u = u(x, t) \in C, x, t \in \mathbb{R}. \quad (1)$$

is the simplest envelope model describing the MI of quasimonochromatic waves in dispersive and weakly nonlinear focusing media. NLS is relevant in water waves (Zakharov 1968; Ablowitz and Segur 1981), in nonlinear optics (Bespalov and Talanov 1966; Solli et al. 2007; Bortolozzo et al. 2007; Pierangeli et al. 2015), in Langmuir waves in plasma (Malomed 2005), in the theory of Bose–Einstein condensates (Bludov et al. 2009;

Pitaevskii and Stringari 2003), etc.. It is well-known that its homogeneous solution

$$u_0(x, t) = a \exp(2i|a|^2 t), \quad (2)$$

$a$  arbitrary complex parameter,

describing Stokes waves (Stokes 1847) in a water wave context, a state of constant light intensity in nonlinear optics, and a state of constant boson density in a Bose–Einstein condensate, is linearly unstable under monochromatic perturbations of wave number  $k$ , if  $0 < |k| < 2|a|$ , implying the existence, in a periodic problem with period  $L$ , of exactly  $N$  unstable modes  $k_j$ ,  $1 \leq j \leq N$ , where

$$N = \left\lfloor \frac{L}{\pi|a|} \right\rfloor, k_j = \frac{2\pi}{L}j, 1 \leq j \leq N, \quad (3)$$

growing exponentially with linear growth rates

$$\sigma_j = k_j \sqrt{4|a|^2 - k_j^2}, 1 \leq j \leq N. \quad (4)$$

Here the term RW is used in an extended sense, and by it we just mean order one (or higher order) coherent structures over the unstable background, generated by MI. A coherent structure of anomalously large amplitude with respect to the average one arises from the rare event in which the nonlinear interaction of many unstable modes is constructive. Therefore the formation of an RW is, strictly speaking, a statistical event. Statistical aspects of the theory of RWs can be found in Gelash et al. (2019), Dematteis et al. (2019), and El and Tobvis (2020); here we concentrate instead on the analytic aspects of the deterministic theory of periodic RWs for a finite number of unstable modes.

The integrable nature of the focusing NLS equation (Zakharov and Shabat 1972) allows one to construct a large family of exact solutions corresponding to perturbations of the background, by degenerating finite-gap solutions (Its et al. 1988), when the spectral curve becomes rational, or using classical Darboux transformations (Matveev and Salle 1991), dressing techniques (Zakharov and Shabat 1974; Zakharov and Gelash 2014), and the Hirota method (Hirota 1976). Among these basic solutions, playing a relevant



role in the theory of RWs, we mention the Peregrine soliton (Peregrine 1983), rationally localized in  $x$  and  $t$  over the background (2), the Kuznetsov (1977)–Ma (1979) soliton, exponentially localized in space over the background and periodic in time, the Akhmediev breather (AB) (Akhmediev et al. 1985), periodic in  $x$  and exponentially localized in time over the background (2), its elliptic generalizations (Akhmediev and Korneev 1986; Akhmediev et al. 1987) and its multi-breather generalizations (Its et al. 1988). Peregrine, Kuznetsov–Ma, and Akhmediev solutions have been observed in experiments (Chabchoub et al. 2011; Kibler et al. 2010, 2012; Kimmoun et al. 2016; Mussot et al. 2018; Pierangeli et al. 2018). Generalizations of these solutions to the case of integrable multi-component NLS equations, characterized by a richer spectral theory, have also been found (Baronio et al. 2012; Degasperis et al. 2018, 2019).

Concerning the NLS Cauchy problem in which the initial condition consists of a generic small perturbation of the exact background (2), what we call the Cauchy problem of the RWs, if such a perturbation is localized, then slowly modulated periodic oscillations described by the elliptic solution of (1) play a relevant role in the longtime regime (Biondini and Kovacic 2014; Biondini et al. 2016). If the initial perturbation is periodic, numerical and real experiments indicate that the solutions of NLS exhibit time recurrence (Yuen and Ferguson 1978; Lake et al. 1977; Yuen and Lake 1982; Akhmediev 2001; Simaey et al. 2001; Kimmoun et al. 2016; Mussot et al. 2018; Pierangeli et al. 2018) and numerically induced chaos (Ablowitz and Herbst 1990; Ablowitz et al. 1993, 2001).

The periodic Cauchy problem of the RWs for the NLS eq. (1) has been solved in Grinevich and Santini (2018a, 2019a), to leading order and in terms of elementary functions, in the case of a finite number  $N$  of unstable modes, and, as we shall see in the next section, it is based on a proper adaptation of the finite-gap method. See also Grinevich and Santini (2018b) for an alternative approach to the study of the RW recurrence, based on matched asymptotic expansions; Grinevich and Santini (2019b) and Grinevich and Santini (2021) for the study of the numerical and linear/nonlinear instabilities of the AB; Grinevich and

Santini (2018c) for the analytic study of the phase resonances in the RW recurrence; and Santini (2018) for the analytic study of the periodic RW recurrence in the PT-symmetric NLS equation (Ablowitz and Musslimani 2013).

## The Finite Gap Method and the Periodic NLS Cauchy Problem of the Rogue Waves, for a Finite Number of Unstable Modes

### Periodic Problem for the Focusing Nonlinear Schrödinger Equation

The integration of the self-focusing NLS equation (1) is based on the zero-curvature representation constructed by Zakharov and Shabat (1972). Indeed it is easy to check that the following pair of linear problems

$$\Psi_x(\lambda, x, t) = \widehat{U}\Psi(\lambda, x, t), \quad (5)$$

$$\Psi_t(\lambda, x, t) = \widehat{V}\Psi(\lambda, x, t), \quad (6)$$

is compatible if and only if the potential  $u$  satisfies (1)  $\forall \lambda$ . Here

$$\begin{aligned} \widehat{U} &= \widehat{U}(\lambda, x, t) = \begin{bmatrix} -i\lambda & iu \\ i\bar{u} & i\lambda \end{bmatrix}, \widehat{V} = \widehat{V}(\lambda, x, t) \\ &= \begin{bmatrix} -2i\lambda^2 + iu\bar{u} & 2i\lambda u - u_x \\ 2i\lambda\bar{u} + \bar{u}_x & 2i\lambda^2 - iu\bar{u} \end{bmatrix}, \end{aligned}$$

and  $\Psi(\lambda, x, t)$  denotes a 2-component vector:

$$\Psi(\lambda, x, t) = \begin{bmatrix} \Psi_1(\lambda, x, t) \\ \Psi_2(\lambda, x, t) \end{bmatrix}.$$

Equation (5) can be rewritten as the following spectral problem

$$\begin{aligned} \mathcal{L}\Psi(\lambda, x, t) &= \lambda\Psi(\lambda, x, t), \text{ where } \mathcal{L} \\ &= \begin{bmatrix} i\partial_x & u \\ -\bar{u} & -i\partial_x \end{bmatrix}. \end{aligned} \quad (7)$$

In contrast to linear PDEs, the methods for solving the soliton systems like NLS admitting zero-curvature representation (or Lax

representation) essentially depend on the boundary conditions. The integration of soliton systems in the class of decaying at large  $x$  potentials is obtained through the spectral transform, introduced in Gardner et al. (1967) to solve the Korteweg–de Vries (KdV) equation, and subsequently used to integrate the NLS equation (Zakharov and Shabat 1972). For each fixed time  $t$ , this transform maps  $u(x, t)$  to the scattering data of the operator  $\mathcal{L}$ , and their time evolution is remarkably linear if  $u(x, t)$  evolves according to NLS.

The spatially periodic (quasiperiodic) problem for soliton systems requires a completely different technique, known as the **finite-gap method**. It was first invented by S.P. Novikov (Novikov 1974) for the spatially periodic KdV problem. Finite-gap solutions of NLS were first constructed by Its and Kotljarov (1976).

Let us define the direct spectral transform for the operator (7). Consider a fixed time  $t = t_0$ , and denote by  $u(x)$  the  $x$ -periodic Cauchy datum  $u(x) = u(x, t_0)$ ,  $u(x + L) = u(x)$ . The direct periodic spectral transform associates with  $u(x)$  the following spectral data:

1. A Riemann surface  $\Gamma$  (the spectral curve)
2. A collection of points on  $\Gamma$  corresponding to the Dirichlet-type spectrum of  $\mathcal{L}$  – the **divisor of zeroes**
1. **The Spectral Curve.** By analogy with solid-state physics, the Bloch functions of  $\mathcal{L}$  are defined as quasiperiodic eigenfunctions of  $\mathcal{L}$  such that

$$\Psi(x + L, t_0) = \kappa \Psi(x, t_0), \text{ where } \kappa = \kappa(\lambda) \in \mathbb{C} \setminus 0. \quad (8)$$

Using the fact that they are eigenfunctions of the  $2 \times 2$  monodromy matrix  $T(\lambda, x_0, t_0)$ , which is holomorphic in the spectral parameter  $\lambda$  in the whole complex plane, one can easily check that the Bloch multipliers  $\kappa$  and the Bloch function  $\Psi(x, t_0)$  are meromorphic on a two-sheeted covering  $\Gamma$  of the  $\lambda$ -plane. To emphasize this fact we write  $\kappa(\gamma)$ ,  $\Psi(\gamma, x, t_0)$ ,  $\gamma \in \Gamma$ . This Riemann surface  $\Gamma$  is called the **spectral curve**. If  $u$  satisfies NLS, then  $\Gamma$  does not depend on  $t_0$ ,

therefore the branch points of  $\Gamma$  provide a complete set of conservation laws for NLS. Let us point out that these conservation laws are strongly nonlocal, and the standard local conservation laws can be obtained by expanding the **quasimomentum**

$$p(\gamma) = \frac{1}{iL} \log(\kappa(\gamma)) \quad (9)$$

near infinity. The function  $p(\gamma)$  is multivalued on  $\Gamma$ , but its differential  $dp(\gamma)$  is well-defined and meromorphic on  $\Gamma$  with two simple poles at the infinity points.

The “classical” spectrum of  $\mathcal{L}$  in  $L^2(\mathbb{R})$  is defined by the condition  $\text{Im } p(\gamma) = 0$ . The branch points of  $\Gamma$  coincide with the ends of the spectral zones. The operator  $\mathcal{L}$  is nonself-adjoint, therefore this spectrum is a collection of arcs in the complex plane, and it may have a very nontrivial structure. If the spectral parameter  $\lambda$  is real, then the matrix  $U$  in (5) is skew-hermitian; therefore the spectrum of  $\mathcal{L}$  contains the whole real line.

Generically, the genus of  $\Gamma$  is infinite. The main idea of Novikov (1974) was to consider a special subclass of **finite-gap potentials** such that the spectral curve  $\Gamma$  is algebraic, that is, it has a finite number of branch points. These potentials form a dense subset in the space of all periodic potentials, and they can be treated as nonlinear analogues of finite Fourier series in the linear theory. Let us remark that the analytic continuation of the Bloch eigenfunctions to the complex domain was first considered by Kohn (1959) for the 1-dimensional Schrödinger operator. The generic infinite-gap case was also studied in the literature (see (Feldman et al. 2003)), but the final formulas in this case are less explicit.

2. **Divisor** Let us fix a point  $x = x_0$ . The auxiliary spectrum of  $\mathcal{L}$  is the set of points  $\gamma_j \in \Gamma$  such that  $\Psi_1(\gamma_j, x_0, t_0) = 0$ , where  $\Psi_1$  denotes the first component of the Bloch eigenfunction. Therefore, it is called the **divisor of zeroes**. It depends on the normalization point  $(x_0, t_0)$ .

Assume that the genus  $g$  of  $\Gamma$  is finite. Then the divisor  $\mathcal{D}(x, t) = \sum_j \gamma_j(x, t)$  contains exactly  $g$  points.

The finite-gap integration essentially uses the following fact: after applying the Abel transform  $\mathbf{A}$  to  $\mathcal{D}(x, t)$ , the  $x$  and  $t$  dynamics become linear; this fact was first established for KdV by Dubrovin (1975) and Its and Matveev (1975).

Let us remark that, in contrast to the case of the real KdV and defocusing NLS equations, the characterization of admissible divisors for the inverse

problem of focusing NLS is nontrivial, and it was obtained by Cherednik (1980).

Therefore, the highly nontrivial spectral transform  $u(x, t) \rightarrow (I, \mathbf{A}(\mathcal{D}(x_0, t)))$  linearizes NLS. The corresponding NLS solution in the finite-gap case can be written explicitly in terms of the Riemann theta-functions ((Its and Kotljarov 1976), see also (Previato 1985)):

$$u(x, t) = C \exp(\mathcal{U}x + \mathcal{V}t) \frac{\theta(\mathbf{A}(\infty_-) - \mathbf{U}_1 x - \mathbf{U}_2 t - \mathbf{A}(\mathcal{D}) - \mathbf{K}|B)}{\theta(\mathbf{A}(\infty_+) - \mathbf{U}_1 x - \mathbf{U}_2 t - \mathbf{A}(\mathcal{D}) - \mathbf{K}|B)}, \quad (10)$$

where  $C$ ,  $\mathcal{U}$ , and  $\mathcal{V}$  are constants defined in terms of the spectral curve;  $\mathbf{A}(\mathcal{D})$ ,  $\mathbf{A}(\infty_+)$ , and  $\mathbf{A}(\infty_-)$  are the Abel transforms of the divisor and of the infinity points of  $\Gamma$  respectively;  $\mathbf{K}$  is the so-called vector of Riemann constants;  $B$  is the Riemann period matrix for  $\Gamma$ ;  $\mathbf{U}_1$  and  $\mathbf{U}_2$  are the vectors of the  $b$ -periods for the quasimomentum and quasi-energy differentials, respectively; and  $\theta(z|B)$  denotes the Riemann theta-function of genus  $g$

$$\theta(z|B) = \sum_{\substack{n_j \in \mathbb{Z} \\ j=1, \dots, g}} \exp \left[ \frac{1}{2} \sum_{j,k=1}^g b_{jk} n_j n_k + \sum_{j=1}^g n_j z_j \right], \quad (11)$$

where  $b_{jk}$  are the components of matrix  $B$ . More information about theta-functional formulas can be found in Dubrovin (1981) and Belokolos et al. (1994). Formula (10) is explicit, but all parameters in it are transcendental expressions in terms of the spectral curve  $\Gamma$ ; therefore, it is rather difficult to use it effectively. Fortunately, as we show below, for a special class of NLS solutions describing the dynamics of RWs generated by  $O(\epsilon)$  perturbations of the unstable background (2), the spectral curve  $\Gamma$  is close to a rational one; therefore, it is possible to write elementary approximate formulas in terms of the Cauchy data. This procedure involves three steps:

- We approximate a generic periodic RW type solution by a 2  $N$ -gap one, where  $N$  is the number of unstable modes.

- For any  $t$ , we approximate the finite-gap solution, up to  $O(\epsilon)$  corrections, by the  $N$ -breather solutions of Akhmediev type; but this approximation depends on the time interval under consideration.
- We keep only “visible” modes; that is, we approximate our solution up to  $O(\epsilon^p)$  error,  $0 < p < 1$ , by  $\mathcal{N}(t)$ -breather solutions,  $\mathcal{N}(t) \leq N$ .

Let us point out that, although the formulas describing the periodic NLS RW dynamics are elementary, they nevertheless take into account the strongly nonlinear character of the NLS equation.

## Finite-Gap Approximation

### Cauchy Problem for the RWs

If the focusing NLS equation is used as a mathematical model for the generation and dynamics of periodic RWs, the initial data are assumed to be small perturbations of the unstable background (2); therefore, we consider the following initial data:

$$u(x, 0) = a[1 + \epsilon v(x)], \quad 0 < \epsilon \ll 1. \quad (12)$$

Since the NLS equation admits the following scaling symmetry

if  $u(x, t)$  solves NLS then  $c u(c x, c^2 t)$  also solves NLS, for any real  $c$

without loss of generality we may assume  $a = 1$ . We work in the spatially periodic setting,

therefore  $v(x)$  is periodic  $v(x+L) = v(x)$ , and it is natural to write it as a Fourier series:

$$v(x) = \sum_{j \geq 1}^{\infty} (c_j e^{ik_j x} + c_{-j} e^{-ik_j x}), k_j = \frac{2\pi}{L} j. \quad (13)$$

The assumption made here that the perturbation  $v(x)$  has zero mean value is not restrictive, due to the scaling symmetry, and it simplifies the calculations. We also assume that the period  $L$  be generic, that is, that  $L/\pi$  is not an integer.

It is natural to write  $\mathcal{L}$  as the sum of the leading term and the perturbation

$$\mathcal{L} = \mathcal{L}_0 + \epsilon \mathcal{L}_1, \mathcal{L}_0 = \begin{bmatrix} i\partial_x & 1 \\ -1 & -i\partial_x \end{bmatrix},$$

$$\mathcal{L}_1 = \begin{bmatrix} 0 & v(x) \\ -\overline{v(x)} & 0 \end{bmatrix},$$

and calculate the spectral data for  $\mathcal{L}$  using the standard perturbation theory near the spectral data for  $\mathcal{L}_0$ .

**The Spectral Data for the Unperturbed Operator**  
The unperturbed spectral curve  $\Gamma_0$  for  $\mathcal{L}_0$  is rational, and a point  $\gamma \in \Gamma_0$  is a pair of complex numbers  $\gamma = (\lambda, \mu)$  satisfying the following quadratic equation:

$$\mu^2 = \lambda^2 + 1.$$

It has two branch points  $\lambda_0 = i$ ,  $\overline{\lambda_0} = -i$  and infinitely many **resonant points** (or double points)  $\lambda_n, -\lambda_n$ , where

$$\mu_n = \frac{\pi n}{L}, \lambda_n = \sqrt{\mu_n^2 - 1}, \operatorname{Re} \lambda_n + \operatorname{Im} \lambda_n > 0,$$

$$n = 1, 2, \dots, \infty,$$

(see Krichever 1977, 1989, 1992). At the resonant points, both Bloch eigenfunctions are either periodic (for even  $n$ ) or antiperiodic (for odd  $n$ ) (Fig. 1).

Since the background (2) is unstable under monochromatic perturbations of wave number  $k_n$  such that  $0 < |k_n| < 2$ , and since  $k_n = 2\mu_n$ , it follows that the background is unstable if the corresponding  $\lambda_n$  are imaginary ( $|\mu_n| < 1$ ) and stable if the  $\lambda_n$  are real ( $|\mu_n| \geq 1$ ). Therefore, we have a finite number  $N = \lfloor \frac{L}{\pi} \rfloor$  of “unstable” resonant points (responsible for the generation of RWs) and an infinite number of “stable” resonant points.

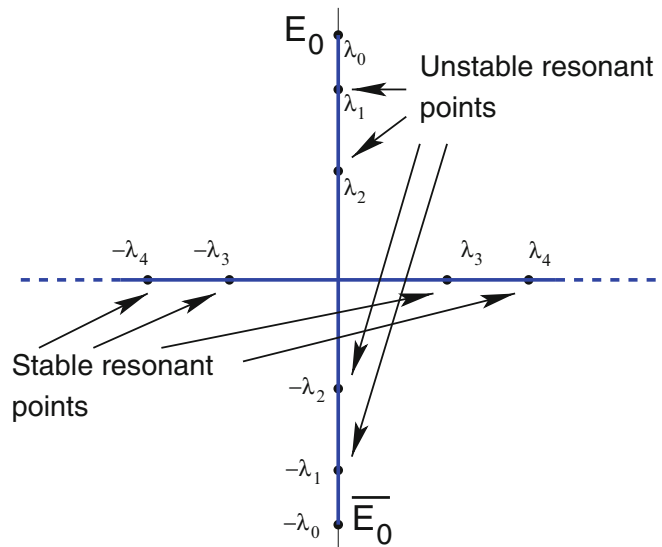
It is convenient to parameterize the unstable resonant points in terms of the angles  $\phi_j$ :

$$\lambda_j = i \sin(\phi_j), \mu_j = \cos(\phi_j), j = 1, \dots, N. \quad (14)$$

### Periodic Rogue Waves and Perturbation Theory, Fig. 1

The spectrum of the unperturbed operator  $\mathcal{L}_0$

### The Spectral Data for the Perturbed Operator



Let us introduce the following notation:

$$\begin{aligned}\alpha_n &= (\mu_n - \lambda_n)\overline{c_n} - (\mu_n + \lambda_n)c_{-n}, \\ \beta_n &= (\mu_n + \lambda_n)\overline{c_{-n}} - (\mu_n - \lambda_n)c_n, \\ \tilde{\alpha}_n &= (\mu_n + \lambda_n)\overline{c_n} - (\mu_n - \lambda_n)c_{-n}, \\ \tilde{\beta}_n &= (\mu_n - \lambda_n)\overline{c_{-n}} - (\mu_n + \lambda_n)c_n.\end{aligned}\quad (15)$$

From a straightforward calculation (see Grinevich and Santini 2018a, 2019a), it follows that the resonant point's  $\lambda_n$  and  $-\lambda_n$  generically split into the pairs of branch points  $\{E_{2n-1}, E_{2n}\}$  and  $\{\tilde{E}_{2n-1}, \tilde{E}_{2n}\}$ :

$$\begin{aligned}E_0 &= i + O(\epsilon^2) \\ E_l &= \lambda_n \mp \frac{\epsilon}{2\lambda_n} \sqrt{\alpha_n \beta_n} + O(\epsilon^2), \\ \tilde{E}_l &= -\lambda_n \pm \frac{\epsilon}{2\lambda_n} \sqrt{\tilde{\alpha}_n \tilde{\beta}_n} + O(\epsilon^2), \\ l &= 2n - 1, 2n.\end{aligned}$$

(see Fig. 2). Here we assume that  $\operatorname{Re} \sqrt{\alpha_n \beta_n} \geq 0$ ,  $\operatorname{Re} \sqrt{\tilde{\alpha}_n \tilde{\beta}_n} \geq 0$  for the unstable points, and  $\operatorname{Im} \sqrt{\alpha_n \beta_n} < 0$ ,  $\operatorname{Im} \sqrt{\tilde{\alpha}_n \tilde{\beta}_n} < 0$  for the stable ones. For the perturbations of the unstable points, we have  $\tilde{E}_l = \overline{E_l}$ .

### Periodic Rogue Waves and Perturbation Theory

**Theory, Fig. 2** The spectrum of the perturbed operator  $\mathcal{L} = \mathcal{L}_0 + \epsilon \mathcal{L}_1$

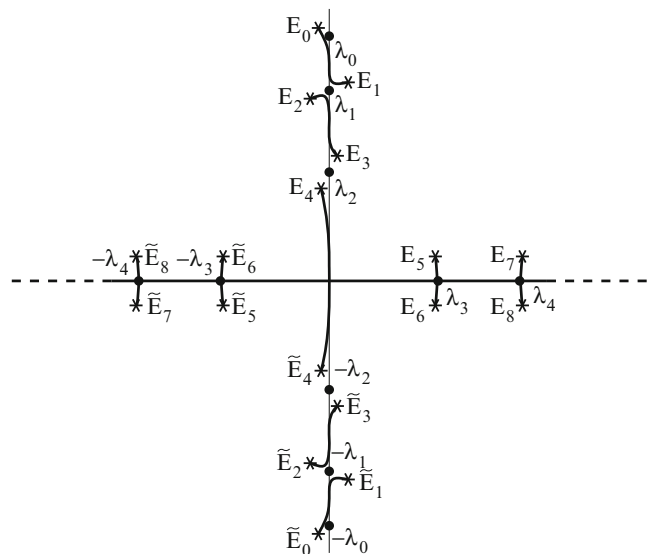
Analogously (Grinevich and Santini 2018a), the divisor points and the quasimomentum at divisor points are given by:

$$\begin{aligned}\lambda(\gamma_n) &= \lambda_n + \frac{\epsilon}{4\lambda_n} [(\mu_n + \lambda_n)\alpha_n + (\mu_n - \lambda_n)\beta_n] + O(\epsilon^2), \\ p(\gamma_n) &= \frac{\epsilon}{4\mu_n} [(\mu_n + \lambda_n)\alpha_n - (\mu_n - \lambda_n)\beta_n] + O(\epsilon^2), \\ \lambda(\gamma_{-n}) &= -\lambda_n - \frac{\epsilon}{4\lambda_n} [(\mu_n - \lambda_n)\tilde{\alpha}_n + (\mu_n + \lambda_n)\tilde{\beta}_n] + O(\epsilon^2), \\ p(\gamma_{-n}) &= \frac{\epsilon}{4\mu_n} [(\mu_n - \lambda_n)\tilde{\alpha}_n - (\mu_n + \lambda_n)\tilde{\beta}_n] + O(\epsilon^2).\end{aligned}\quad (16)$$

### The Leading Order Finite-Gap Solution

A generic small perturbation of the constant solution generates an infinite genus spectral curve, but the perturbations corresponding to stable resonant points remain of order  $\epsilon$  for all  $t$  and can be well-described by the linear perturbation theory. Therefore one can close all gaps associated with stable points, obtaining a finite-gap approximation of the spectral curve. We also keep only divisor points located near unstable resonant points. Let us remark that this finite-gap approximation is rather nonstandard.

For such spectral curves, all parameters in formulas (10) and (11) can be calculated to leading order in terms of elementary functions of the Fourier coefficients of  $v(x)$ . Let us introduce the following notation:



$$\begin{aligned}\widehat{\lambda}_j &= \lambda_j, \widehat{\mu}_j = \mu_j, \widehat{\gamma}_j = \gamma_j, \widehat{\phi}_j = \phi_j, \\ \widehat{\lambda}_{j+N} &= -\lambda_j, \widehat{\mu}_{j+N} = \mu_j, \widehat{\gamma}_{j+N} = \gamma_{-j}, \\ \widehat{\phi}_{j+N} &= -\phi_j, j = 1, \dots, N.\end{aligned}$$

The leading order solution of the Cauchy problem has the following form (Grinevich and Santini 2019a):

$$u(x, t) = \exp(2it) \frac{\theta(\mathbf{z}_+(x, t)|B)}{\theta(\mathbf{z}_-(x, t)|B)} \times (1 + O(\epsilon)), \quad (17)$$

where

$$\begin{aligned}\theta(z|B) &= \sum_{n_j} \exp \left[ 2\pi i \sum_j n_j z_j + \pi i \sum_{j,k} b_{j,k} n_j n_k \right], \\ j, k &= 1, \dots, 2N,\end{aligned}$$

$$\begin{aligned}b_{jj} &= \frac{1}{\pi i} \log \left[ \frac{\epsilon \sqrt{\alpha_j \beta_j}}{4 \sin(2\phi_j) \cos(\phi_j)} \right] \\ &\quad + O(\epsilon), \\ b_{j+N, j+N} &= -\overline{b_{jj}}, 1 \leq j \leq N, \\ b_{kj} &= \frac{1}{\pi i} \log \left| \frac{\sin\left(\frac{\widehat{\phi}_j - \widehat{\phi}_k}{2}\right)}{\cos\left(\frac{\widehat{\phi}_j + \widehat{\phi}_k}{2}\right)} \right| \\ &\quad + O(\epsilon^2) \text{ for all } j \neq k\end{aligned} \quad (18)$$

(for  $\epsilon = 0$  formula (18) was derived in Its et al. (1988),

$$\begin{aligned}\mathbf{z}_{\pm}(x, t) &= \mp \mathbf{A}(\infty_-) - \mathbf{U}_1 x - \mathbf{U}_2 t - \mathbf{A}(D), \\ \mathbf{A}(\infty_-) &= \left[ -\frac{1}{4} - \frac{\phi_1}{2\pi}, \dots, -\frac{1}{4} - \frac{\phi_N}{2\pi} \mid -\frac{1}{4} + \frac{\phi_1}{2\pi}, \dots, -\frac{1}{4} + \frac{\phi_N}{2\pi} \right] + O(\epsilon^2), \\ \mathbf{A}(D) &= \frac{1}{2\pi i} \left[ \log \left( \frac{e^{i\phi_1} \alpha_1}{\sqrt{\alpha_1 \beta_1}} \right), \dots, \log \left( \frac{e^{i\phi_1} \alpha_N}{\sqrt{\alpha_N \beta_N}} \right) \mid \log \left( \frac{e^{-i\phi_1} \overline{\beta_1}}{\sqrt{\alpha_1 \beta_1}} \right), \dots, \log \left( \frac{e^{-i\phi_N} \overline{\beta_N}}{\sqrt{\alpha_N \beta_N}} \right) \right] + \dots, \\ \mathbf{U}_1 &= \frac{1}{\pi} [-\cos(\phi_1), \dots, -\cos(\phi_N) \mid -\cos(\phi_1), \dots, -\cos(\phi_N)] + O(\epsilon^2), \\ \mathbf{U}_2 &= \frac{1}{\pi} [-\sin(2\phi_1), \dots, -\sin(2\phi_N) \mid \sin(2\phi_1), \dots, \sin(2\phi_N)] + O(\epsilon^2).\end{aligned} \quad (19)$$

### The Solution of the Cauchy Problem in Terms of Elementary Functions

Since formula (17) provides the solution up to  $O(\epsilon)$  corrections, it is enough to sum the exponents in the theta-function over the elementary hypercube in  $\mathbb{R}^{2N}$  containing the trajectory point  $-\mathbf{w}(t)$  (far away vertices contribute with smaller negligible terms), where

$$\mathbf{w}(t) = (\text{Im} B)^{-1} \text{Im}(\mathbf{z}_{\pm}(x, t)).$$

To avoid big arguments as  $t$  increases, it is convenient to shift the arguments to the basic

elementary cell, using the periodicity properties of the theta function (Dubrovin 1981):

$$(\widetilde{\mathbf{z}}_{\pm})_j(x, t) = (\mathbf{z}_{\pm}(x, t))_j - \sum_k b_{jk} [w_j(t)],$$

obtaining

$$\begin{aligned}u(x, t) &= \exp(2it + 2i\Phi) \frac{\widetilde{\theta}(\widetilde{\mathbf{z}}_+(x, t)|B)}{\widetilde{\theta}(\widetilde{\mathbf{z}}_-(x, t)|B)} \\ &\quad \times (1 + O(\epsilon)),\end{aligned} \quad (20)$$

where



$$\Phi = \sum_{j=1}^{2N} \left( \frac{\pi}{2} + \widehat{\phi}_j \right) [w_j(t)],$$

$$\widetilde{\theta}(z|B) = \sum_{\substack{n_j \in \{-1,0\} \\ j=1,2,\dots,2N}} \exp \left[ \pi i \left( \sum_{l=1}^{2N} \sum_{s=1}^{2N} b_{ls} n_l n_s + 2 \sum_{l=1}^{2N} n_l z_l \right) \right]. \quad (21)$$

In (21) we have a finite sum of elementary functions, instead of an infinite series, and (20) represents the  $N$ -breather solution of Akhmediev type (Its et al. 1988). Therefore, for each  $t$ , our finite-gap solution is approxi-

mated by such solutions, but the parameters of the  $N$ -breathers are different in different time intervals. Let us remark that the boundaries of these time intervals are calculated explicitly in terms of the initial data (Grinevich and Santini 2019a).

### Keeping Only Visible Modes

Shifting the hypercube by  $1/2$  in all directions ( $\widehat{n}_j = 2n_j + 1$ ), one can write:

$$u(x, t) = \exp(2it + 2i\Phi) \frac{\widehat{\theta}(\widehat{z}_+(x, t)|B)}{\widehat{\theta}(\widehat{z}_-(x, t)|B)} (1 + O(\epsilon)),$$

$$\widehat{\theta}(\widehat{z}|B) = \sum_{\substack{\widehat{n}_j \in \{-1,1\} \\ j=1,2,\dots,2N}} \exp \left[ \sum_{\substack{l,s=1 \\ l \neq s}}^{2N} \log \left| \frac{\sin \left( \frac{\widehat{\phi}_l - \widehat{\phi}_s}{2} \right)}{\cos \left( \frac{\widehat{\phi}_l + \widehat{\phi}_s}{2} \right)} \right| \frac{\widehat{n}_l \widehat{n}_s}{4} + \sum_{l=1}^{2N} \frac{\widehat{n}_l \widehat{z}_l}{2} - \frac{1}{2} \sum_{l=1}^{2N} \widehat{z}_l \right], \quad (22)$$

where

$$\widehat{z}_j = \widetilde{z}_j - \frac{1}{2} \sum_k b_{jk}.$$

Let us denote

$$\widetilde{w}_j(t) = w_j(t) - [w_j(t)].$$

To obtain an  $O(\epsilon^p)$ ,  $0 < p < 1$  approximation of the solution, we use the following rule:

1. If  $\widetilde{w}_j < \frac{1-p}{2}$ , we keep only the terms with  $n_j = 1$  in (22).
2. If  $\frac{1-p}{2} \leq \widetilde{w}_j \leq \frac{1+p}{2}$ , we keep the terms with  $n_j = 1$  and with  $n_j = -1$ .
3. If  $\widetilde{w}_j > \frac{1+p}{2}$ , we keep only the terms with  $n_j = -1$ .

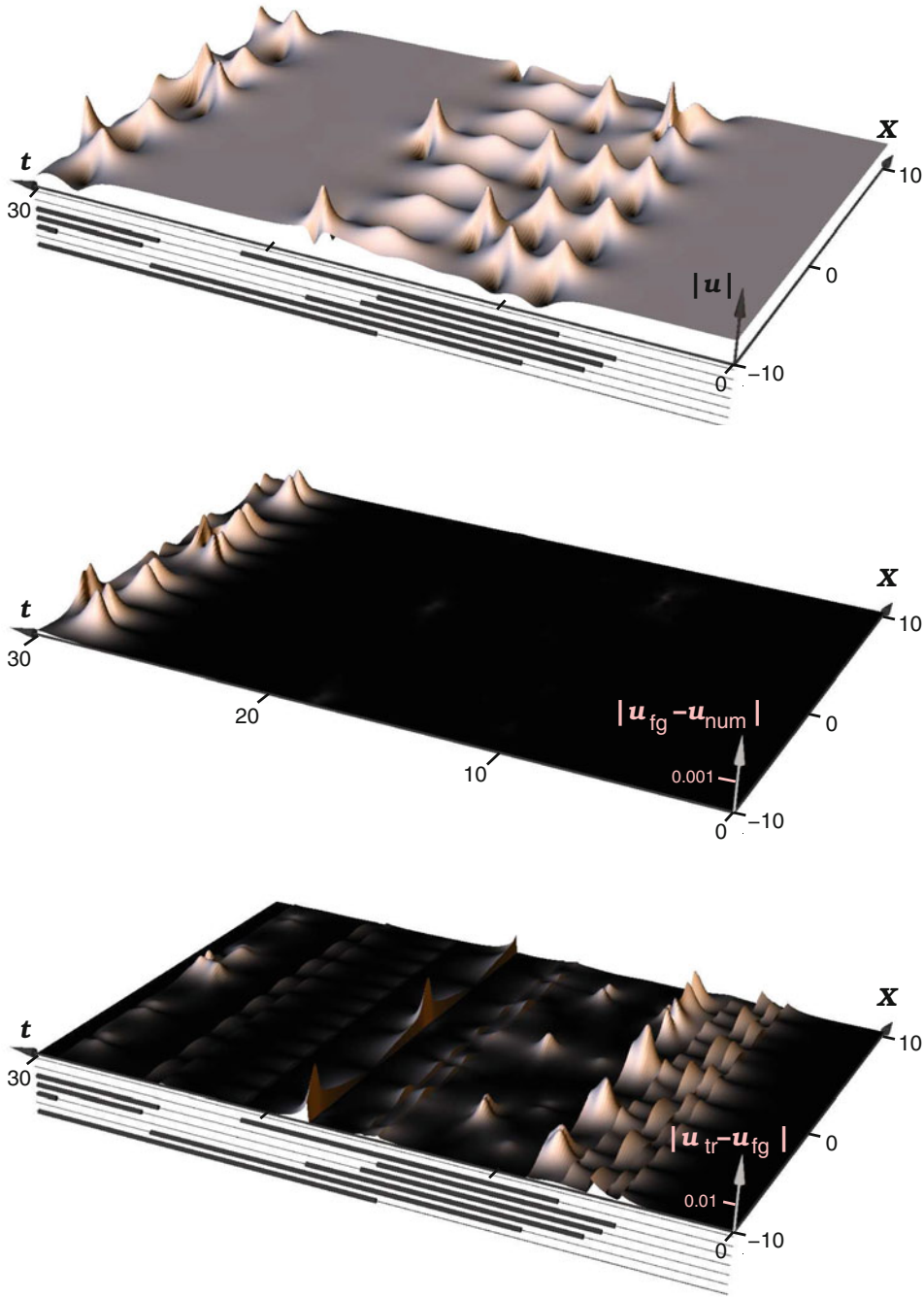
Therefore, for each generic  $t$ , we have a summation over the vertices of a hypercube of

dimension  $2\mathcal{N}(t)$ ,  $\mathcal{N}(t) \leq N$ . In our approximation, the off-diagonal terms of matrix  $B$  do not depend on  $\epsilon$ ; therefore, in each time interval  $I$  of the partition generated by the initial data, the approximation function does not depend on  $\epsilon$ , and the solution of the Cauchy problem is approximated by an **exact** NLS solution, the  $\mathcal{N}(I)$ -breather solution of Akhmediev type.

Let us show a numerical simulation illustrating these three steps. We take  $L = 20$  (6 unstable modes),  $0 \leq t \leq 30$ , and consider a perturbation of  $u_0(x) \equiv 1$  with the following parameters:

$$\begin{aligned} c_1 &= 0.5, c_{-1} = 0.3 + 0.3i, \\ c_2 &= 0.5, c_{-2} = -0.03 + 0.03i, \\ c_3 &= 0.3, c_{-3} = 0.2 + 0.3i, c_4 = 0.3, \\ c_{-4} &= -0.3 + 0.03i, \\ c_5 &= 0.3, c_{-5} = 0.2 + 0.3i, \\ c_6 &= 0.3, c_{-6} = -0.3 + 0.03i, \epsilon = 10^{-6}. \end{aligned} \quad (23)$$

We assume  $p = 1/2$  (Fig. 3).



**Periodic Rogue Waves and Perturbation Theory, Fig. 3** The upper graph shows the graph of  $|u(x, t)|$  obtained using the analytic formula (20), coinciding pixel-to-pixel with numerical simulation. The middle graph shows the absolute value of the difference between the numerical solution and the full-hypercube finite-gap approximation,

multiplied by  $10^3$ . The difference at the left-hand side of the picture, of order  $10^{-3}$ , is very likely to be a numerical artifact. The graph below shows the absolute value of the difference between the full-hypercube finite-gap approximation and the approximation involving the relevant vertices only, multiplied by  $10^2$ ; its magnitude is  $O(10^{-2})$ , a little

### The Case $N = 1$ and the Fermi–Pasta–Ulam–Tsingou Recurrence of RWs

Hereafter, it is convenient to reintroduce the amplitude  $a$  of the background. If  $\pi/|a| < L < 2\pi/|a|$ , then  $N = 1$  and the solution is well approximated by a genus 2 exact solution on a Riemann surface with  $O(\epsilon)$  handles, expressible, to leading order, in terms of elementary functions of the initial data. Let

$$\begin{aligned}\alpha_1 &= e^{-i\phi_1} \bar{c}_1 - e^{i\phi_1} c_{-1}, \\ \beta_1 &= e^{i\phi_1} \bar{c}_{-1} - e^{-i\phi_1} c_1,\end{aligned}\quad (24)$$

then the solution of the Cauchy problem to leading order (up to  $O(\epsilon)$  corrections), in the finite interval  $0 \leq t \leq T$ , reads as follows (Grinevich and Santini 2018a, b):

$$\begin{aligned}u(x, t) &= \sum_{m=0}^n \mathcal{A}\left(x, t; \phi_1, x^{(m)}, t^{(m)}\right) e^{i\rho^{(m)}} \\ &\quad - \frac{1 - e^{4in\phi_1}}{1 - e^{4i\phi_1}} a e^{2i|a|^2 t}, x \in [0, L],\end{aligned}\quad (25)$$

where the parameters  $x^{(m)}$ ,  $t^{(m)}$ ,  $\rho^{(m)}$ , and  $m \geq 0$  are defined as:

$$\begin{aligned}x^{(m)} &= X^{(1)} + (m-1)\Delta X, t^{(m)} = T^{(1)} + (m-1)\Delta T, \\ X^{(1)} &= \frac{\arg \alpha_1}{k_1} + \frac{L}{4}, \Delta X = \frac{\arg(\alpha_1 \beta_1)}{k_1}, (\text{mod } L), \\ T^{(1)} &\equiv \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^2}{2|a|^4 \epsilon |\alpha_1|} \right), \\ \Delta T &= \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^4}{4|a|^8 \epsilon^2 |\alpha_1 \beta_1|} \right), \\ \rho^{(m)} &= 2\phi_1 + (m-1)4\phi_1, n = \left\lceil \frac{T - T^{(1)}}{\Delta T} + \frac{1}{2} \right\rceil,\end{aligned}\quad (26)$$

and function  $\mathcal{A}$  is the AB solution:

$$\begin{aligned}\mathcal{A}(x, t; \theta, X, T) &:= \\ &:= a e^{2i|a|^2 t} \frac{\cos[\sigma(\theta)(t-T) + 2i\theta] + \sin \theta \cos[k(\theta)(x-X)]}{\cosh[\sigma(\theta)(t-T)] - \sin \theta \cos[k(\theta)(x-X)]}, \\ k(\theta) &= 2|a| \cos \theta, \quad \sigma(\theta) = k(\theta) \sqrt{4|a|^2 - k^2(\theta)} = 2|a|^2 \sin(2\theta),\end{aligned}\quad (27)$$

exact solution of NLS for all real parameters  $\theta$ ,  $X$ , and  $T$ .

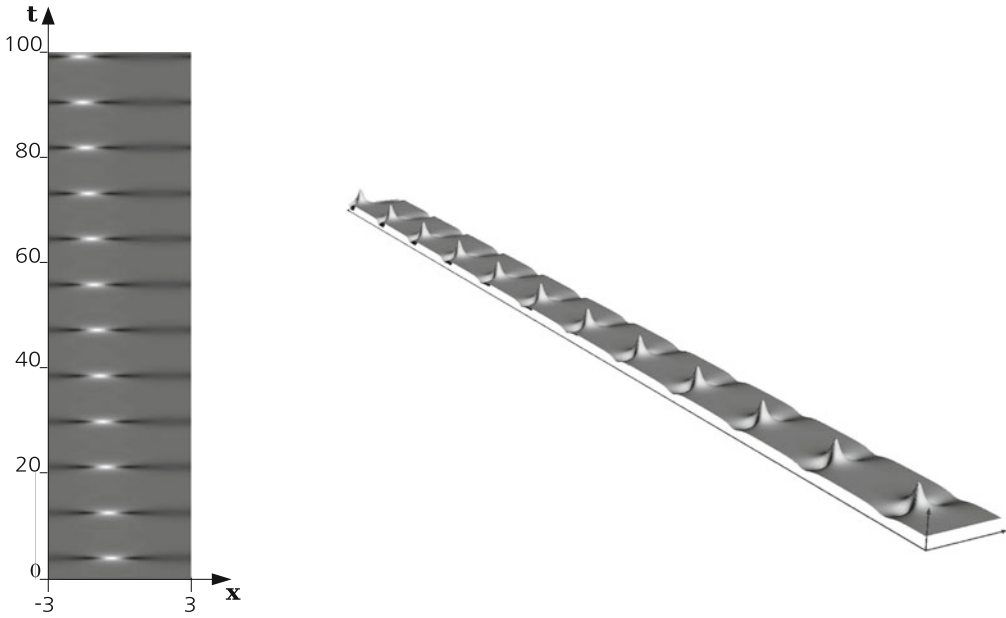
The solution (24)–(27) shows an exact recurrence of RWs described by the AB, whose parameters change at each appearance according to (26). It gives the quantitative description of the Fermi–Pasta–Ulam–Tsingou (FPUT) recurrence (Gallavotti 2008), without thermalization, of the NLS RWs in terms of elementary functions.  $X^{(1)}$  and  $T^{(1)}$  are respectively the position and the time of the first appearance;  $\Delta X$  is the  $x$ -shift of the position of the RW between two consecutive

appearances, and  $\Delta T$  is the recurrence time (the time between two consecutive appearances) (see Fig. 4). Therefore,  $T^{(1)}$  and  $\Delta T$  are the characteristic times of the RW recurrence.

We remark that the RW recurrence is described, to leading order, by just the four real parameters  $X^{(1)}$ ,  $T^{(1)}$ ,  $\Delta X$ , and  $\Delta T$ . Four free real parameters appear, indeed, in the unstable part of the initial condition (the real and imaginary parts of  $c_1$  and  $c_{-1}$ ). We also remark that formulas (24)–(27), in perfect quantitative agreement with the output of the corresponding numerical experiments (Grinevich

**Periodic Rogue Waves and Perturbation Theory, Fig. 3** (continued) higher than  $\epsilon^{1/2}$  but, taking into account that the full hypercube contains  $4^6 = 4096$  points, the agreement is sufficiently good. The horizontal lines in the upper and lower pictures show which modes are “visible”;

that is, for which modes we keep both terms with  $n_j = 1$  and  $n_j = -1$  in the summation. Hereafter brighter and darker colors in the plots correspond to higher and lower values of  $|u(x, t)|$ , respectively. See Grinevich and Santini (2019a) for more details.



**Periodic Rogue Waves and Perturbation Theory, Fig. 4** Density and 3D plots of  $|u(x, t)|$  with  $-L/2 \leq x \leq L/2$ ,  $0 \leq t \leq 100$ ,  $L = 6$ ,  $\epsilon = 10^{-4}$ ,  $a = 1$ , with a generic initial condition  $c_{-1} = 0.3 + 0.3i$ ,  $c_1 = 0.5$ , obtained

using the refined split-step method (Javanainen and Ruostekoski 2006), in extremely good quantitative agreement with (25) and (26)

and Santini 2018a), were successfully tested in nonlinear optics experiments involving a photorefractive crystal (Pierangeli et al. 2018) and optical fibers (Naveau et al. 2019).

We end this section remarking that from (16),

$$(E_1 - E_2)^2 = -\epsilon^2 \frac{|a|^2}{\sin^2 \phi_1} \alpha_1 \beta_1, \quad (28)$$

it is possible to express the FPUT recurrence period  $\Delta t$  and  $x$ -shift  $\Delta x$  in terms of the unstable gap:

$$\begin{aligned} \Delta T &= \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^4}{4|a|^6 \sin^2 \phi_1 |E_1 - E_2|^2} \right), \\ \Delta X &= \frac{\arg(-(E_1 - E_2)^2)}{k_1}. \end{aligned} \quad (29)$$

## RW Perturbation Theory

Like other model equations in physics, NLS provides only a first-order description of reality and,

in order to increase the agreement between theoretical predictions and experimental results, additional terms  $f[u]$  must be added to the equation:

$$u_t - i(u_{xx} + 2|u|^2 u) = f[u], \quad (30)$$

and perturbation theory can be used if these terms are small to describe their effect on the RW dynamics. For instance, dissipation can hardly be avoided in all natural phenomena involving RWs, and it is often described by terms of the type  $f[u] = v_j |u|^{2j} u$ ,  $v_j < 0$ ,  $j \geq 0$ ,  $|v_j| \ll 1$ . Then the background solution and the linear growth rate of (30) read, up to  $O(v_j^2)$ ,

$$\begin{aligned} \tilde{u}_0(x, t, v) &= a e^{2i|a|^2 t} \exp \left[ v_j |a|^{2j} t (1 + 2i|a|^2 t) \right], \\ \tilde{\sigma}(t, v) &= k \sqrt{4|a|^2 \exp \left( 2v_j |a|^{2j} t \right) - k^2}. \end{aligned} \quad (31)$$

If the initial perturbation is sufficiently small, a small dissipation can quench the growth process before the nonlinear effects become relevant,

stabilizing the MI (Segur et al. 2005). In any case, due to (31), instability is always canceled if the time interval of interest is sufficiently long (Segur et al. 2005; Kharif et al. 2009). Therefore, the presence of dissipation introduces another characteristic time  $T_{diss} = (1/|v_j|) \log(2|a|/k)$ , the time at which the unstable mode  $k$  becomes stable, and the stabilizing effect takes place when  $T_{diss} < T^{(1)}$ . But what happens in the interesting case in which dissipation is small and  $T_{diss} \gg T^{(1)}$ ?

A few interesting results in this direction are known. An experiment in a tank showing that the AB initial condition evolves into a recurrence of ABs whose position is shifted by  $\Delta X = L/2$ , and a numerical experiment showing that the same AB initial condition evolves, according to the NLS equation perturbed by a small linear dissipation, into the same pattern, thus interpreting the result of the tank experiment as the effect of dissipation (Kimmoun et al. 2016). A real sinusoidal initial perturbation of the background, evolving numerically according to focusing NLS perturbed by linear loss or gain terms, gives rise to a recurrence of ABs with shifts respectively  $\Delta X = L/2$  or  $\Delta X = 0$  (Soto-Crespo et al. 2017).

The proper analytic model describing quantitatively and in terms of elementary functions the  $O(1)$  effects of a small perturbation on the dynamics of periodic NLS AWs, providing, in particular, the explanation of the above real and numerical experiments, was constructed and applied to the case of a linear loss or gain in Coppini et al. (2020), and it was applied to the physically relevant complex Ginzburg-Landau (CGL) (Newell and Whitehead 1969) and Lugiato-Lefever (1987) models, treated as perturbations of NLS, in Coppini and Santini (2020). The physical mechanism for the  $O(1)$  effects generated by the small perturbation

terms is explained qualitatively as follows. In the finite time interval when the RW appears, the nonintegrable perturbations generate a small correction to the unperturbed solution, becoming an  $O(1)$  correction when the next RW appears, due to MI.

If  $[0, T]$  is the time interval in which one studies the dynamics, hereafter we assume that

$$|v_j|T, |v_j||a|^{2(j+1)}T^2 \ll 1, \quad (32)$$

implying from (31) that one can neglect the slow decay/growth of the amplitude and of the oscillation frequency of the background. Then  $a$  can be treated as a constant during the dynamics.

If  $u$  evolves according to NLS, the spectral curve and the trace of the monodromy matrix are constants of motion. Now we calculate how they evolve in time in the presence of a small perturbation  $f[u]$ . A variation  $\delta U$  of the matrix potential  $U = \begin{bmatrix} 0 & u \\ \bar{u} & 0 \end{bmatrix}$  induces the following variation of the monodromy matrix

$$\delta(\text{tr } T)(\lambda, t) = i \int_0^L \text{tr} \left[ \hat{T}(\lambda, x + L, x, t) \delta U(x, t) \right] dx, \quad (33)$$

where  $\hat{T}(\lambda, x, y, t)$  is the fundamental matrix solution of (5) such that  $\hat{T}(\lambda, y, y, t)$  is the identity matrix, connected to the monodromy matrix  $T$  through  $T(\lambda, t) = \hat{T}(\lambda, L, 0, t)$ .

Choosing  $\delta U = U_t dt = [i\sigma_3(U_{xx} + 2U^3) + F] dt$ , where  $F = \begin{bmatrix} 0 & f \\ \bar{f} & 0 \end{bmatrix}$ , then  $\delta \text{tr } T = (\text{tr } T)_t dt$ , and recalling that the NLS vector field does not change  $\text{tr } T$ , one obtains the variation of  $\text{tr } T$  in the time interval  $[0, t]$ :

$$\Delta(\text{tr } T)(\lambda_1, t) = i \int_0^t \int_0^L dx \left[ \hat{T}_{21}(\lambda_1, x + L, x, \tilde{t}) f(x, \tilde{t}) + \hat{T}_{12}(\lambda_1, x + L, x, \tilde{t}) \overline{f(x, \tilde{t})} \right], \quad (34)$$

where to leading order (Coppini et al. 2020):

$$\Delta(\text{tr } T)(\lambda_1, t) = \frac{\sin^2(\phi_1)L^4}{4\pi^2} \Delta\left((E_1 - E_2)^2\right). \quad (35)$$

The complicated integral (34) can be evaluated in terms of elementary functions to leading order. Near each RW appearance, the solution is well approximated by the AB, and far from the RW

appearance, the integral over the  $x$ -period is exponentially small in  $t$ . Therefore, the integral over the finite time interval of each RW appearance can be well approximated by the integral of the AB over the whole line  $t \in \mathbb{R}$ , and, to leading order,

$$\Delta \text{tr } T(\lambda_1, t) = n_{\text{app}} J(\lambda_1),$$

where  $n_{\text{app}}$  is the number of RW appearances in the time interval  $[0, t]$ ,

$$J(\lambda_1) = i \int_{-\infty}^{+\infty} dt \int_0^L dx \left[ \widehat{T}_{21}(\lambda_1, x + L, x, t) f(x, t) + \widehat{T}_{12}(\lambda_1, x + L, x, t) \overline{f(x, t)} \right], \quad (36)$$

and  $u, \widehat{T}$  correspond to the AB and can be evaluated in terms of elementary functions using the Darboux transformations of NLS (Yurov and Yurov 2018) as follows (Coppini et al. 2020):

$$u(x, t) = ae^{2i|a|^2t} \frac{\cosh[\sigma_1 t + 2i\phi_1] - \sin(\phi_1) \sin(k_1 x)}{\cosh(\sigma_1 t) + \sin(\phi_1) \sin(k_1 x)} \quad (37)$$

coincides with the AB (27) up to translation symmetries, and

$$\begin{aligned} \widehat{T}(\lambda_1, x + L, x, t) &= -I + \frac{2\pi i \lambda_1^2}{\mu_1^3} \frac{g(x, t)}{|q_1|^2 + |q_1|^2} \begin{bmatrix} -\overline{q_2 q_1} & -\overline{q_2 q_2} \\ \overline{q_1 q_1} & \overline{q_1 q_2} \end{bmatrix}, \\ g(x, t) &:= ae^{2i|a|^2t} q_2^2 - \overline{a} e^{-2i|a|^2t} q_1^2 - 2i|a| \sin \phi_1 q_1 q_2, \end{aligned} \quad (38)$$

where  $I$  is the identity matrix, and

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = 2 \begin{bmatrix} ae^{i|a|^2t} \cos\left(\frac{k_1}{2}x - \phi_1/2 + i\frac{\sigma_1}{2}t\right) \\ i|a|e^{-i|a|^2t} \sin\left(\frac{k_1}{2}x + \phi_1/2 + i\frac{\sigma_1}{2}t\right) \end{bmatrix}. \quad (39)$$

If the perturbation consists of a linear loss or gain  $f = v_0 u$ ,  $|v_0| \ll 1$  ( $v_0 > 0$  in the case of a gain, and  $v_0 < 0$  in the case of a loss), then  $J(\lambda_1) = \frac{\pi^2 \sin \phi_1}{|a|^2 \cos^3 \phi_1}$ , and (35) gives the following variation of the gap due to a single AW appearance

$$\Delta_1((E_1 - E_2)^2) = 4v \cot(\phi_1), \quad (40)$$

with

$$(E_1 - E_2)^2|_{t=0} = -\frac{\epsilon^2 |a|^2 \alpha_1 \beta_1}{\sin^2 \phi_1}, \quad (41)$$

implying the following discrete dynamics of the gap

$$\left(E_1^{(m)} - E_2^{(m)}\right)^2 = -\frac{\epsilon^2 |a|^2 \alpha_1 \beta_1}{\sin^2 \phi_1} + 4mv \cot \phi_1, \quad m \geq 0, \quad (42)$$

where  $E_1^{(m)}$  and  $E_2^{(m)}$  are the positions of the branch points of the gap after the  $m$ th RW appearance (see Fig. 6). At last, introducing the sequence  $\{Q_m\}$  such that

$$Q_m = -\left(\frac{\sin^2(\phi_1)}{\epsilon|a|}\right)^2 \left(E_1^{(m)} - E_2^{(m)}\right)^2, \quad (43)$$

then



$$Q_m = \alpha_1 \beta_1 - \frac{v_0}{\epsilon^2} \frac{\sigma_1}{|a|^4} m, m \geq 0. \quad (44)$$

Using now the relation (29) between the spectral gap and the parameters of the FPUT recurrence, we infer that a generic periodic perturbation of the background evolves according to the perturbed NLS equation

$$u_t - i(u_{xx} + 2|u|^2 u) = v_0 u \quad (45)$$

into the RW recurrence of ABs (25), with essentially the same  $x^{(1)}$ ,  $t^{(1)}$  as in (26), but now

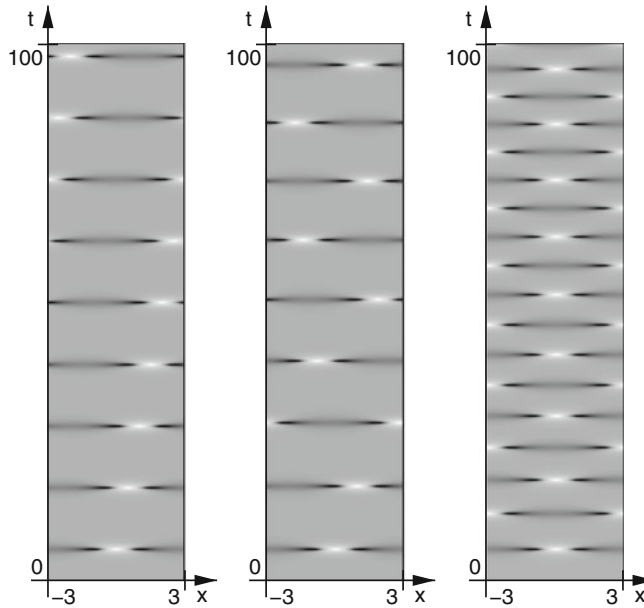
$$\begin{aligned} \Delta X_m := x^{(m+1)} - x^{(m)} &= \frac{\arg(Q_m)}{k_1} \pmod{L}, \\ \Delta T_m := t^{(m+1)} - t^{(m)} &= \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^4}{4|a|^8 \epsilon^2 |Q_m|} \right), \end{aligned} \quad (46)$$

where  $Q_m$  is defined in (44),  $\alpha_1, \beta_1$  are defined in (24), and  $\rho^{(m)}$  in (26).

From formulas (25), (44), and (46), we distinguish the following cases (assuming that  $|\alpha_1 \beta_1|, \sigma_1, |a| = O(1)$ ).

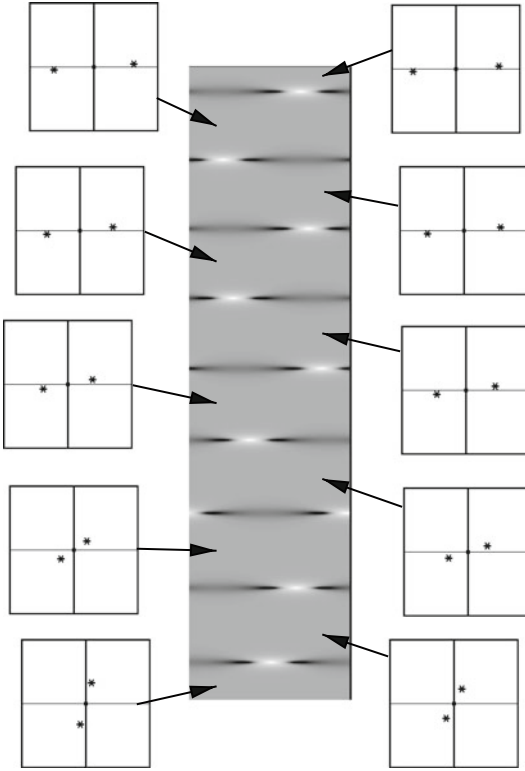
1. If  $|v_0| \ll \epsilon^2$ , then  $Q_m \sim \alpha_1 \beta_1$  for every  $m$ , and there is no basic difference with the zero-gain/loss case.
2. If  $|v_0|$  is  $O(\epsilon^2)$ , we have a transient, consisting of a few RW recurrences, in which  $Q_m \rightarrow -(v_0/\epsilon^2)(\sigma_1/|a|^4)m$  as  $m$  increases, and the solution tends to one of the two asymptotic states characterized by the following elementary formulas (see the central picture of Fig. 5):

$$\begin{aligned} \arg Q_m = \pi &\Rightarrow \Delta X_m = L/2, \text{ if } v_0 > 0 \text{ (loss),} \\ \arg Q_m = 0 &\Rightarrow \Delta X_m = 0, \text{ if } v_0 < 0 \text{ (gain),} \\ \Delta T_m &= \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^3}{4|a|^4 |v_0| m} \right). \end{aligned} \quad (47)$$



**Periodic Rogue Waves and Perturbation Theory, Fig. 5** The density plot of  $|u(x, t)|$  for generic initial data:  $c_1 = 0.5$  and  $c_{-1} = 0.15 - 0.2i$ . Hereafter, we assume  $-L/2 \leq x \leq L/2, L = 6, 0 \leq t \leq 100, \epsilon = 10^{-4}, a = 1$ . From left to right:  $v_0 = 0$ ,  $v_0 = 10^{-9} < \epsilon^2 = 10^{-8}$ , and  $v_0 = 10^{-5} \gg \epsilon^2$ . In the left figure, we have the RW recurrence described by formulas (25)–(26). In the central figure, the solution tends to the SVA (47) with  $\Delta X_m \rightarrow L/2$ , after a relatively

long transient. In the right figure, after the first appearance, the solution enters, without any transient, the SVA (47) with  $\Delta X_m = L/2, m \geq 1$ . Hereafter the first appearance is almost the same for different  $v_0$ . These numerical integrations are in extremely good agreement with the above theoretical formulas: differences are smaller than the theoretically expected error  $O(\epsilon)$ . See Coppini et al. (2020) for more details.



**Periodic Rogue Waves and Perturbation Theory, Fig. 6** illustrates the time evolution of the gap  $E_1 - E_2$ , due to each RW appearance in the central picture of Fig. 5. See Coppini et al. (2020) for more details.

3. If  $|v_0| \gg \epsilon^2$ , then  $Q_m \sim -(v_0/\epsilon^2)(\sigma_1/|a|^4)m$ ,  $m \geq 1$ , and, after the first AW appearance (essentially not affected by loss/gain) and without any transient, the solution enters immediately one of the above two recurrence patterns (47) for  $m \geq 1$  (see the right picture of Fig. 5).

These two asymptotic states describe *lower-dimensional* recurrence patterns depending on just two real parameters defining their position in space-time, unlike the zero-loss/gain case. In addition, while the difference between two consecutive recurrence times is finite:

$$|\Delta T_{m+1} - \Delta T_m| = \frac{1}{\sigma_1} \log \left( \frac{m+1}{m} \right), \quad (48)$$

the relative difference  $|\Delta T_{m+1} - \Delta T_m|/\Delta T_m$  is small, since  $v_0$  is small. Therefore, we have slowly varying lower-dimensional asymptotic states that

can be viewed as slowly varying attractors (SVAs), completely ruled by the parameter  $v_0$ .

If we apply the above theory to the CGL equation

$$\begin{aligned} u_t - i(u_{xx} + 2|u|^2 u) &= f[u], u = u(x, t) \in \mathbb{C}, \\ f[u] &:= \gamma u_{xx} + v_0 u + v_1 |u|^2 u + v_2 |u|^4 u, \end{aligned} \quad (49)$$

a well-known applicable envelope equation describing the propagation of a small amplitude quasimonochromatic pulse in a nonlinear and dispersive medium, in the presence of diffusion and linear and nonlinear loss or gain, where  $\gamma > 0$  and  $v_j \in \mathbb{R}, j = 0, 1, 2$ , and we treat (49) as perturbation of NLS:  $\gamma, |v_j| \ll 1, j = 0, 1, 2$ , we have again the above FPUT recurrence of ABs, but now (Coppini and Santini 2020)

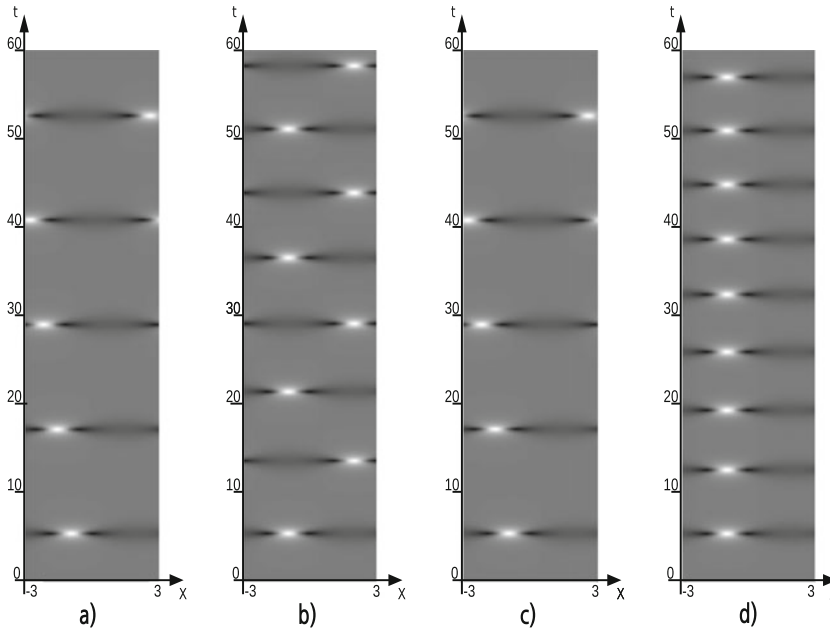
$$\begin{aligned} Q_m &= \alpha_1 \beta_1 + \epsilon^{-2} \mathcal{P}_{CGL} m, m \geq 0, \\ \mathcal{P}_{CGL} &:= \frac{\sigma_1}{|a|^2} \left\{ -\frac{8}{3} \gamma \sin^2 \phi_1 + \frac{v_0}{|a|^2} + \frac{v_1}{3} [10 - 7 \cos(2\phi_1)] \right. \\ &\quad \left. + v_2 \frac{|a|^2}{15} [218 - 234 \cos(2\phi_1) + 31 \cos(4\phi_1)] \right\}. \end{aligned} \quad (50)$$

As before, if  $|\mathcal{P}_{CGL}| \ll \epsilon^2$ , then  $Q_m \sim \alpha_1 \beta_1$  for every  $m$ , and there is no basic difference with the unperturbed NLS case. If  $|\mathcal{P}_{CGL}| = O(\epsilon^2)$ , we have a transient in which  $Q_m \rightarrow \epsilon^{-2} \mathcal{P}_{CGL} m$  as  $m$  increases, and the solution tends to one of the two asymptotic states characterized by the following elementary formulas:

$$\begin{aligned} \arg Q_m = \pi &\Rightarrow \Delta X_m = L/2, \text{ if } \mathcal{P}_{CGL} < 0 \text{ (global loss),} \\ \arg Q_m = 0 &\Rightarrow \Delta X_m = 0, \text{ if } \mathcal{P}_{CGL} > 0 \text{ (global gain),} \\ \Delta T_m &= \frac{1}{\sigma_1} \log \left( \frac{\sigma_1^4}{4|a|^8 |\mathcal{P}_{CGL}| m} \right). \end{aligned} \quad (51)$$

If  $|\mathcal{P}_{CGL}| \gg \epsilon^2$ , then  $Q_m \sim \epsilon^{-2} \mathcal{P}_{CGL} m, m \geq 1$ , and, after the first RW appearance and without any transient, the solution enters one of the above two recurrence patterns (51) for  $m \geq 1$  (see Fig. 7b, d).

An important special case is when diffusion, loss, and gain balance ( $\mathcal{P}_{CGL} = 0$ ):



**Periodic Rogue Waves and Perturbation Theory, Fig. 7** The density plot of  $|u(x, t)|$  describing the FPUT recurrence for generic initial data:  $c_1 = -0.16 + 0.41i$  and  $c_{-1} = 0.55 - 0.11i$ , for  $0 \leq t \leq 60$ ,  $-L/2 \leq x \leq L/2$ ,  $L = 6$ ,  $\epsilon = 10^{-4}$ ,  $a = 1$ . **a)** pure NLS ( $v_0 = v_1 = v_2 = \gamma = 0$ ); **b)** CGL with global loss ( $v_0 = -10^{-6}$ ,  $v_1 = 10^{-6}$ ,  $v_2 = -10^{-7}$ ,  $\gamma = 10^{-6} \Rightarrow \mathcal{P}_{CGL} = -1.04 \cdot 10^{-6}$ ); **c)** CGL in which loss and gain exactly balance ( $v_0 = -10^{-8}$ ,  $v_1 =$

$1.133099 \cdot 10^{-8}$ ,  $v_2 = -10^{-9}$ ,  $\gamma = 10^{-8} \Rightarrow \mathcal{P}_{CGL} = 0$ ), coinciding pixel-to-pixel with the pure NLS dynamics; **d)** CGL with global gain ( $v_0 = -10^{-6}$ ,  $v_1 = 2 \cdot 10^{-6}$ ,  $v_2 = -10^{-7}$ ,  $\gamma = 10^{-6} \Rightarrow \mathcal{P}_{CGL} = 6.79 \cdot 10^{-6}$ ). These numerical integrations are in extremely good agreement with the above theoretical formulas: differences are smaller than the theoretically expected error  $O(\epsilon)$ . See Coppini and Santini (2020) for more details.

$$\begin{aligned}
 & -\frac{8}{3}\gamma \sin^2 \phi_1 + \frac{v_0}{|a|^2} + \frac{v_1}{3} [10 - 7 \cos(2\phi_1)] \\
 & + v_2 \frac{|a|^2}{15} [218 - 234 \cos(2\phi_1) + 31 \cos(4\phi_1)] = 0.
 \end{aligned}
 \tag{52}$$

In this case, the effect of the perturbation on the NLS RW dynamics is zero to leading order, and the NLS FPUT recurrence of RWs described in (24)–(27) stabilizes for several recurrences (see Fig. 7c). Eq. (52) is a linear constraint on the four coefficients of diffusion and loss or gain, and can be satisfied in many possible ways, balancing, i.e., diffusion with linear or nonlinear gain, or balancing linear loss with nonlinear gain.

## Future Directions

The analytic theory developed to study the dynamics of NLS periodic RWs opens a certain

number of research directions including: (1) The study of the theory when the number  $N$  of unstable modes is large; (2) the generalization of the above results to other integrable and applicable NLS type equations: (i) multicomponent NLS equations, like the relativistic Thirring model (Thirring 1958) and the Manakov system (Manakov 1976); (ii) discrete NLS equations, like the Ablowitz-Ladik lattice (Ablowitz and Ladik 1975) (partial results in some of these directions can be found in Coppini (2021)); (3) the generalization of these results to the Davey–Stewartson equations (Davey and Stewartson 1974), integrable and applicable 2+1 dimensional generalizations of the NLS equation, to study the RW dynamics in multidimensions. The illustrated RW perturbation theory opens the following research directions: (1) the relaxation of the restrictions (32), taking account of the slow changes of the amplitude and phase of the background; (2) the generalization of the theory to the

case of several unstable modes; and (3) the study of the effect of Hamiltonian perturbations, like the quintic Hamiltonian correction to NLS.

## References

- Ablowitz MJ, Herbst B (1990) On homoclinic structure and numerically induced chaos for the nonlinear Schrödinger equation. *SIAM J Appl Math* 50(2): 339–351
- Ablowitz MJ, Ladik JF (1975) Nonlinear differential-difference equations. *J Math Phys* 16(3):598–603
- Ablowitz MJ, Musslimani ZH (2013) Integrable nonlocal nonlinear Schrödinger equation. *Phys Rev Lett* 110(6):064105
- Ablowitz MJ, Segur H (1981) Solitons and the inverse scattering transform. *SIAM studies in applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia
- Ablowitz MJ, Schober CM, Herbst BM (1993) Numerical chaos, roundoff errors and homoclinic manifolds. *Phys Rev Lett* 71(17):2683–2686
- Ablowitz MJ, Hammack J, Henderson D, Schober CM (2001) Long-time dynamics of the modulational instability of deep water waves. *Phys D* 152:416–433
- Akhmediev N, Dudley JM, Solli DR, Turitsyn SK (2013) Recent progress in investigating optical rogue waves. *J Opt* 15(6):060201
- Akhmediev NN (2001) Nonlinear physics: Déjà vu in optics. *Nature (London)* 413:267–268
- Akhmediev NN, Korneev VI (1986) Modulation instability and periodic solutions of the nonlinear Schrödinger equation. *Theor Math Phys* 69(2):1089–1093
- Akhmediev NN, Eleonskii VM, Kulagin NE (1985) Generation of periodic trains of picosecond pulses in an optical fiber: exact solutions. *Sov. Phys. JETP* 62(5): 894–899
- Akhmediev NN, Eleonskii VM, Kulagin NE (1987) Exact first order solutions of the nonlinear Schrödinger equation. *Theor Math Phys* 72(2):809–818
- Bailung H, Sharma SK, Nakamura Y (2011) Observation of Peregrine solitons in a multicomponent plasma with negative ions. *Phys Rev Lett* 107:255005
- Baronio F, Degasperis A, Conforti M, Wabnitz S (2012) Solutions of the vector nonlinear Schrödinger equations: evidence for deterministic rogue waves. *Phys Rev Lett* 109(2):44102
- Belokolos ED, Bobenko AI, Enolski VZ, Its AR, Matveev VB (1994) *Algebro-geometric approach in the theory of Integrable equations*, Springer series in nonlinear dynamics. Springer, Berlin
- Benjamin TB, Feir JE (1967) The disintegration of wave trains on deep water. Part I. theory. *J Fluid Mech* 27(3): 417–430
- Bespalov VI, Talanov VI (1966) Filamentary structure of light beams in nonlinear liquids. *JETP Lett* 3(12): 307–310
- Biondini G, Kovacic G (2014) Inverse scattering transform for the focusing nonlinear Schrödinger equation with nonzero boundary conditions. *J Math Phys* 55(3):031506
- Biondini G, Li S, Mantzavinos D (2016) Oscillation structure of localized perturbations in modulationally unstable media. *Phys Rev E* 94(6):060201(R)
- Bludov YV, Konotop VV, Akhmediev N (2009) Matter rogue waves. *Phys Rev A* 80(3):033610
- Bortolozzo U, Montina A, Arecchi FT, Huignard JP, Residori S (2007) Spatiotemporal pulses in a liquid crystal optical oscillator. *Phys Rev Lett* 99(2):023901
- Chabchoub A, Hoffmann NP, Akhmediev N (2011) Rogue wave observation in a water wave tank. *Phys Rev Lett* 106(20):204502
- Cherednik IV (1980) Reality conditions in “finite-zone integration”. *Sov Phys Dokl* 25:450–452
- Coppini F (2021) Towards the theory of anomalous waves in nature and nonlinear Schrödinger type equations. PhD thesis, Università di Roma, La Sapienza
- Coppini F, Santini PM (2020) The Fermi-Pasta-Ulam-Tsingou recurrence of periodic anomalous waves in the complex Ginzburg-Landau and in the Lugiato-Lefever equations. *Phys Rev E* 102:062207
- Coppini F, Grinevich PG, Santini PM (2020) The effect of a small loss or gain in the periodic NLS anomalous wave dynamics. I. *Phys Rev E* 101:032204
- Davey A, Stewartson K (1974) On three-dimensional packets of surface waves. *Proc R Soc Lond A* 338: 101–110
- Degasperis A, Lombardo S, Sommacal M (2018) Integrability and linear stability of nonlinear waves. *J Nonlinear Sci* 28(4):1251–1291
- Degasperis A, Lombardo S, Sommacal M (2019) Rogue wave type solutions and spectra of coupled nonlinear Schrödinger equations. *Fluids* 4:57
- Dematteis G, Grafke T, Onorato M, Vanden-Eijnden E (2019) Experimental evidence of hydrodynamic instantons: the universal route to rogue waves. *Phys. Rev. X* 9:041057
- Dubrovin BA (1975) Inverse problem for periodic finite-zone potentials in the theory of scattering. *Funct. Anal. Appl.* 9(1):61–62
- Dubrovin BA (1981) Theta functions and non-linear equations. *Russ Math Surv* 36(2):11–92
- Dysthe KB, Trulsen K (1999) Note on breather type solutions of the NLS as models for freak-waves. *Phys Scr* T82:48–52
- El G, Tobvis A (2020) Spectral theory of soliton and breather gases for the focusing nonlinear Schrödinger equation. *Phys Rev E* 101:052207
- Feldman J, Knorrer H, Trubowitz E (2003) *Riemann surfaces of infinite genus*, volume 20 of RM monograph series. Am. Math. Soc, Providence
- Gallavotti G (ed) (2008) *The Fermi-Pasta-Ulam problem: a status report*, volume 728 of lecture notes in physics. Springer, Berlin/Heidelberg
- Ganshin AN, Efimov VB, Kolmakov GV, Mezhev-Deglin LP, McClintock PVE (2008) Observation of an inverse

- energy Cascade in developed acoustic turbulence in superfluid helium. *Phys Rev Lett* 101:065303
- Gardner CS, Greene JM, Kruskal MD, Miura RM (1967) Method for solving the Korteweg-de Vries equation. *Phys Rev Lett* 19:1095–1097
- Gelash A, Agafontsev D, Zakharov V, El G, Randoux S, Suret P (2019) Bound state soliton gas dynamics underlying the spontaneous modulational instability. *Phys Rev Lett* 123:234102
- Grinevich PG, Santini PM (2018a) The finite gap method and the analytic description of the exact rogue wave recurrence in the periodic NLS Cauchy problem. 1. Nonlinearity 31(11):5258–5308
- Grinevich PG, Santini PM (2018b) The exact rogue wave recurrence in the NLS periodic setting via matched asymptotic expansions, for 1 and 2 unstable modes. *Phys Lett A* 382(14):973–979
- Grinevich PG, Santini PM (2018c) Phase resonances of the NLS rogue wave recurrence in the quasi-symmetric case. *Theor Math Phys* 196(3):1294–1306
- Grinevich PG, Santini PM (2019a) The finite-gap method and the periodic NLS Cauchy problem of anomalous waves for a finite number of unstable modes. *Russ Math Surv* 74(2):211–263
- Grinevich PG, Santini PM (2019b) Numerical instability of the Akhmediev breather and a finite gap model of it. In: Buchstaber VM et al (eds) Recent developments in integrable systems and related topics of mathematical physics. Springer, pp 3–23. E-print arXiv:1708.00762
- Grinevich PG, Santini PM (2021) The linear and nonlinear instability of the Akhmediev breather. *Nonlinearity* 34(12):8331–8358
- Haver S (1995) Freak wave event at Draupner jacket January 1. (Report). Statoil, Tech. Rep. PTT-KU-MA. Retrieved 2015-06-03, 1995
- Henderson KL, Peregrine DH, Dold JW (1999) Unsteady water wave modulations: fully nonlinear solutions and comparison with the nonlinear Schrödinger equation. *Wave Motion*
- Hirota R (1976) Direct methods for finding exact solutions of nonlinear evolution equations. *Lecture Notes in Mathematics*, Springer/New York, pp 157–176
- Its AR, Kotljarov VP (1976) Explicit formulas for solutions of a nonlinear Schrödinger equation. *Dokl Akad Nauk Ukrain SSR Ser A* 1051:965–968
- Its AR, Matveev VB (1975) Hill's operator with finitely many gaps. *Funct Anal Appl* 9(1):65–66
- Its AR, Rybin AV, Sall MA (1988) Exact integration of nonlinear Schrödinger equation. *Theor Math Phys* 74(1):20–32
- Javanainen J, Ruostekoski J (2006) Symbolic calculation in development of algorithms: split-step methods for the Gross-Pitaevskii equation. *J Phys A* 39(12):L179–L184
- Kharif C, Pelinovsky E (2011) Focusing of nonlinear wave groups in deep water. *JETP Lett* 73:170–175
- Kharif C, Pelinovsky E, Slunyaev A (2009) *Rogue waves in the ocean*. Springer, Berlin, Heidelberg
- Kibler B, Fatome J, Finot C, Millot G, Dias F, Genty G, Akhmediev N, Dudley J (2010) The Peregrine soliton in nonlinear fibre optics. *Nat Phys* 6(10):790–795
- Kibler B, Fatome J, Finot C, Millot G, Genty G, Wetzel B, Akhmediev N, Diaz F, Dudley J (2012) Observation of Kuznetsov-Ma soliton dynamics in optical fibre. *Sci Rep* 2:463
- Kimmoun O, Hsu HC, Branger H, Li MS, Chen YY, Kharif C, Onorato M, Kelleher EJR, Kibler B, Akhmediev N, Chabchoub A (2016) Modulation instability and phase-shifted Fermi-Pasta-Ulam recurrence. *Sci Rep* 6:28516
- Kohn W (1959) Analytic properties of Bloch waves and Wannier functions. *Phys Rev* 115:809–821
- Krichever IM (1977) Methods of algebraic geometry in the theory on nonlinear equations. *Russ Math Surv* 32(6):185–213
- Krichever IM (1989) Spectral theory of two-dimensional periodic operators and its applications. *Russ Math Surv* 44(2):145–225
- Krichever IM (1992) Perturbation theory in periodic problems for two-dimensional integrable systems. *Sov Sci Rev, Sect C, Math Phys Rev* 9(2):1–103
- Kuznetsov EA (1977) Solitons in a parametrically unstable plasma. *Sov Phys Dokl* 22:507–508
- Lake BM, Yuen HC, Rungaldier H, Ferguson WE (1977) Nonlinear deep-water waves: theory and experiment. Part 2. Evolution of a continuous wave train. *J Fluid Mech* (1):83, 49–74
- Liu C, van der Wel REC, Rotenberg N, Kuipers L, Krauss TF, Falco AD, Fratallocchi A (2015) Triggering extreme events at the nanoscale in photonic seas. *Nat Phys* 11(4):358–363
- Lugiato LA, Lefever R (1987) Spatial dissipative structures in passive optical systems. *Phys. Rev. Letters* 85:2209–2211
- Ma YC (1979) The perturbed plane wave solutions of the cubic Schrödinger equation. *Stud Appl Math* 60(1):43–58
- Malomed B (2005) *Nonlinear Schrödinger Equations*, in Scott, Alwyn (ed.), *Encyclopedia of Nonlinear Science*, New York: Routledge, pp. 639–643
- Manakov SV (1976) Note on the integration of Euler's equations of the dynamics of an n-dimensional rigid body. *Funct. Anal. Appl.* 10(4):328–329
- Matveev VB, Salle MA (1991) *Darboux transformations and solitons*, Springer series in nonlinear dynamics. Springer, Berlin/Heidelberg
- Moslem WM, Sabry R, El-Labany SK, Shukla PK (2011) Dust-acoustic rogue waves in a nonextensive plasma. *Phys Rev E* 84:066402
- Mussot A, Naveau C, Conforti M, Kudlinski A, Szriftgiser P, Copie F, Trillo S (2018) Fibre multiwave-mixing combs reveal the broken symmetry of Fermi-pasta-Ulam recurrence. *Nat Photonics* 12(5):303–308
- Naveau C, Szriftgiser P, Kudlinski A, Conforti M, Trillo S, Mussot A (2019) Experimental characterization of

- recurrences and separatrix crossing in modulational instability. *Opt Lett* 44:5426
- Newell AC, Whitehead JA (1969) Review of the finite bandwidth concept. *Proc IUTAM Symposium on Instability of Continuous Systems* 44(Springer-Verlag, Berlin):284–289
- Novikov SP (1974) The periodic problem for the Korteweg-de Vries equation. *Funct Anal Appl* 8(3):236–246
- Onorato M, Residori S, Bortolozzo U, Montina A, Arecchi F (2013) Rogue waves and their generating mechanisms in different physical contexts. *Phys Rep* 528(2): 47–89
- Osborne A, Onorato M, Serio M (2000) The nonlinear dynamics of rogue waves and holes in deep-water gravity wave trains. *Phys Lett A* 275(5–6):386–393
- Peregrine DH (1983) Water waves, nonlinear Schrödinger equations and their solutions. *J Austral Math Soc Ser B* 25:16–43
- Pierangeli D, DiMei F, Conti C, Agranat AJ, DelRe E (2015) Spatial rogue waves in photorefractive ferro-electrics. *Phys Rev Lett* 115(9):093901
- Pierangeli D, Flammini M, Zhang L, Marcucci G, Agranat AJ, Grinevich PG, Santini PM, Conti C, DelRe E (2018) Observation of exact Fermi-Pasta-Ulam-Tsingou recurrence and its exact dynamics. *Phys Rev X* 8(4):041017
- Pitaevskii LP, Stringari S (2003) *Bose-Einstein Condensation*. Clarendon, Oxford
- Previato E (1985) Hyperelliptic quasi-periodic and soliton solutions of the nonlinear Schrödinger equation. *Duke Math J* 52(2):329–377
- Santini PM (2018) The periodic Cauchy problem for PT-symmetric NLS, I: the first appearance of rogue waves, regular behavior or blow up at finite times. *J Phys A Math Theor* 51(49):495207
- Segur H, Henderson D, Carter J, Hammack J, Li C-M, Pheiff D, Socha K (2005) Stabilizing the Benjamin-Feir instability. *J Fluid Mech* 539:229–271
- Simaey G, Emplit P, Haelterman M (2001) Experimental demonstration of the Fermi-Pasta-Ulam recurrence in a modulationally unstable optical wave. *Phys Rev Lett* 87(3):033902
- Solli DR, Ropers C, Koonath P, Jalali B (2007) Optical rogue waves. *Nature* 450:1054–1057
- Soto-Crespo JM, Devine N, Akhmediev N (2017) Adiabatic transformation of continuous waves into trains of pulses. *Phys Rev A* 96:023825
- Stokes G (1847) On the theory of oscillatory waves. *Trans Camb Philos Soc VIII*:197–229. And Supplement 314–326
- Thirring WE (1958) A soluble relativistic field theory. *Ann Phys* 3:91–112
- Yuen H, Lake B (1982) Nonlinear dynamics of deep-water gravity waves. *Adv Appl Mech* 22:67–229
- Yuen HC, Ferguson WE (1978) Relationship between Benjamin-Feir instability and recurrence in the nonlinear Schrödinger equation. *Phys Fluids* 21(8): 1275–1278
- Yurov AV, Yurov VA (2018) The Landau-Lifshitz equation, the NLS, and the magnetic rogue wave as a by-product of two colliding regular “positons”. *Symmetry* 10:82
- Zakharov V, Ostrovsky L (2009) Modulation instability: the beginning. *Phys D Nonlinear Phenomena* 238(5): 540–548
- Zakharov VE (1968) Stability of period waves of finite amplitude on surface of a deep fluid. *J Appl Mech Tech Phys* 9(2):190–194
- Zakharov VE, Gelash AA (2014) Superregular solitonic solutions: a novel scenario for the nonlinear stage of modulation instability. *Nonlinearity* 27(4):R1–R39
- Zakharov VE, Shabat AB (1972) Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov Phys JETP* 34(1):62–69
- Zakharov VE, Shabat AB (1974) A scheme for integrating the nonlinear equations of mathematical physics by the method of the inverse scattering transform I. *Funct Anal Appl* 8(3):226–235



---

# Index

## A

Abelian Lie algebra, 196  
Action-angle coordinates, 307, 313  
Action-angle variable, 34, 156, 249, 292, 438  
Action levels  
    2–1-choreography symmetry, 373  
    Lagrange solutions in the 3-body problem, 378  
    line symmetry, 373  
    mountain pass solution in the 3-body problem, 378  
Adiabatic decoupling, 409, 426, 427, 429  
Adiabatic evolution, 422  
Adiabatic Hamiltonian, 422  
Adiabatic invariant, 300  
Adiabatic limit, 424  
Adiabatic parameter, 411, 420, 421, 426  
Adiabatic perturbation theory, 409  
Adiabatic subspace, 412  
Adiabatic theorems  
    for many-body quantum systems, 429  
    for nonlinear dynamics, 429  
    for open quantum systems, 429  
    for resonances, 428–429  
    without gap condition, 428  
Adjoint  
    action, 84  
    operator, 82  
Affine structure, 31  
Agmon metric, 391, 401–402  
Algebra, of vector fields, 196  
Almost-invariant subspace, 414  
Almost-periodic function, 168, 175  
Analyticity, 461–462  
Analytic maps, 223–224  
Annihilation operator, 472–475, 478, 479, 481  
Anomalous waves, 566  
Anti-Stokes line, 396  
Antisymmetric wave function, 469  
Anti-symmetry, icosahedral group, 370  
Approximate symmetries, 193  
Arbitrary factor, 54  
Area-preserving maps, 28, 283  
Aristotelian physics, 1

Arnold diffusion, 20, 157, 292  
Arnold resonance tongues, 28  
Arnold's diffusion, 284  
Arnold's scheme, 260–261  
    Arnold's transformation, 262–263  
    description, 261–262  
    estimates and convergence, 263–266  
Astronomy, 1  
Asymptotic freedom, 4  
Attracting heteroclinic, 171  
Attracting tori, 171  
Autonomous systems, dynamical systems, 188  
Autoparametric excitation, 180–181  
Autoparametric resonance, 180  
Autoparametric systems, 180  
Averaging, 169  
    principle, 293  
    theorem, 99  
    theory, 262

## B

Baker–Campbell–Hausdorff formula, 191  
Banach spaces, 256, 269, 275  
Benjamin–Feir resonances, 559  
Bernoulli condition, 542  
Berry's connection and parallel transport, 423–424  
Beta function, 4  
Bifibration, 316  
Bifoliation, 316  
Bifurcation, 15, 80  
    curves, 150  
    diagram, 23, 39  
    Hamiltonian Hopf, 445  
    Hopf, 80, 98  
    Hopf–Neimark–Sacker, 80, 98  
    imperfect, 441–443  
    organization of, 442–444  
    quantum, 433  
    quantum bifurcation diagrams, 448–449  
    quasi-periodic Hamiltonian Hopf, 98  
    quasi-periodic Hopf, 98

- Bifurcation (*cont.*)  
   saddle-node, 442  
   and symmetry, 440–441  
   two degree-of-freedom integrable systems, diagrams for, 444–448  
 Biholomorphic map, 223  
 Birkhoff normal form, 230  
   procedure, 557–558  
 Bloch's method, 59–60  
 Blow-up technique, 531  
 3-Body problem, 370  
   mountain pass solutions, 376  
   planar, 372  
   planar symmetry groups, 372  
 Bogoliubov theory, 468, 478  
 Bogoliubov transformation, 476, 478–480, 482  
   and quasi-free states, 473–476  
 Bohr–Sommerfeld rule, 399  
 Border-collision bifurcation, 521, 531  
 Borel, E., 96  
 Borel property, 97  
 Borel summability, 10  
 Borel theorem, 97  
 Born–Oppenheimer approximation, 409, 412–413, 428  
 Born series, 71–73  
 Bose–Einstein condensate, 469, 470, 475, 483  
 Boson, 469–471, 476–477, 483  
 Bosonic Bogoliubov map, 473  
 Bosonization, 480, 482, 484  
 Bosons, 458, 460  
 Boundary singularity, 379  
 Bourgain's approach, 270–271  
 Brillouin–Wigner method, 61–62  
 Broer–Huiteima–Takens (BHT) non-degeneracy condition, 34  
 Bruno's method, 115  
 Bruno vectors, 136  
  
**C**  
 Cancellation, 5  
 Canonical anticommutator relations (CAR), 472, 473  
 Canonical commutator relations (CCR), 472, 473, 475, 481, 482  
 Canonical coordinates, 232, 233  
 Canonical transformation, 116, 153  
 Cantor dust, 15  
 Cantor family, 15  
 Cantorized' singularity, 98  
 Cantor set, 15, 41, 98, 135  
 Cantor stratification, 15  
 Casimirs, 308  
 Cauchy estimate, 257, 280  
 Cauchy problem, 224–226, 546  
 Cayley formula, 8, 462  
 Center, 25  
 Center manifold, 195  
 Center-saddle bifurcation, 22  
 Centralizer, 112, 220  
 Change, of variables, 224  
 Chaos, 15, 21, 25, 29, 37, 39, 41, 42, 181  
 Chaotic attractors, 171  
 Chaotic motions, 1, 5  
   particles, 359  
 Characteristic exponents, 172  
 Chenciner–Montgomery symmetry group, 374  
 Choreography  
    $n$ -body problem, 357, 359  
   non-planar, 369  
   symmetry, 373  
 2–1-Choreography symmetry, 373  
 Circle diffeomorphisms, 25  
 Circle maps, 25, 26  
 Circular restricted three-body problem (RTBP), 158  
 Classical Darboux transformations, 566  
 Classical detuning, 142  
 Classical KAM theory, 31  
 Classical perturbation theory, 340  
   classical theory, 341–342  
   computation of precession of the perihelion, 344–345  
   Delaunay action-angle variables, 342–343  
   perturbing function, expansion of, 344  
   precession of the perihelion of mercury, 342–345  
   restricted, planar, circular, three-body problem, 343–344  
 Cluster, 10  
 Codimension-one M-singularity, 509–510  
 Coexistence, 167, 174, 177  
 Coherent state, 475, 476  
 Collisions  
   asymptotic estimates, 380  
   between two or more particles, 359  
   between two or more particles, 379  
   instants, isolatedness, 380  
 Compass-gait biped, 519  
   definition, 519  
   foot scuffing, 533  
   perturbation theory for Poincaré maps, 528–531  
   Poincaré map of, 520  
 Complete integrability, 113  
 Complex Jordan canonical form, 213  
 Complex linearization, 29  
 Compton scattering, 3  
 Computational methods in perturbation theory, 154  
   conformally symplectic systems, 160  
   equilibrium point of Hamiltonian system, 156–158  
   future research, 160  
   parametrization method, 159–160  
   quasi-periodic motions and KAM theory, 158–159  
   Solar system, 155–156  
   time dependent perturbations, 158  
 Condensed matter, 457, 459, 460  
 Condition  $\omega^\#$ , 113  
 Condition A, 110  
 Conditionally periodic, 16  
 Conformally symplectic systems, 160  
 Conical intersection, 413  
 Conjugacy class, 223

Conjugated flow, 349  
 Conjugation function, 122  
 Conservation laws,  $n$ -body problem, 381  
 Conservation of momentum, 555  
 Conservative perturbations, 24  
 Conserved quantities, 193  
 Constant  
     frequency integrable system, 300  
     of motion, 193  
 Continuous linearization, 83  
 Convergence, 106, 187, 203  
     and problems, 108  
 Convexity, 307  
 Core, symmetry groups, 369  
 Correlation energy, 476, 478, 480  
 Coulomb interaction, 480, 484  
 Coupling constant, 469  
 Covering space, 94  
 Crandall-Rabinowitz bifurcation theorem, 545  
 Creation operator, 471–473, 481, 482  
 Critical energy, 150  
 Critical point universality, 1  
 Cubic polynomial, 144  
 Cyclic actions, symmetry groups, 361  
 Cyclic group  $C_{6k}$ , 368  
 Cyclic group of order  $2n$ , 365

## D

Darboux theorem, 309  
 Darboux transformations, 578  
 Deep sea water, 566  
 Degenerate case, 56  
     Bloch's method, 59–60  
     energy and eigenvectors, corrections to, 57–59  
 Degenerate perturbation theory, 347  
 Degree of freedom, 15  
 Delaunay action-angle variables, 342–343  
 Delaunay elements, 320  
 $\lambda$ -Dependent operator, 59  
 Determining equations, 188  
 Detuned resonant 2-DOF systems, 145  
 Detuning, 142  
     classical, 142  
     parameter  $\delta$ , 142  
 Diagonalization, 475, 478  
 Diagrammatic rule, 124  
     node factor, 125  
     propagator, 125  
 Diffeomorphism, 223, 251  
 Differential equations  
     symmetry of, 186  
     system, 357  
 Differential geometry, 83  
 Dihedral action, symmetry groups, 361  
 Dihedral group  $G=D_{2n}$ , 364  
 4-Dihedral periodic minimizers, 371  
 6-Dihedral symmetric periodic minimizers, 371  
 Diophantine Cantor set, 16, 38, 42

Diophantine condition, 15, 95, 97, 547  
     Bryuno condition, 127  
     Melnikov condition, 135  
     standard, 127  
 Diophantine frequencies, 307  
 Diophantine frequency vector, 15  
 Diophantine property, 8, 10, 11  
 Diophantine vector, 250, 295  
 Dirichlet boundary condition, 233, 401  
 Dirichlet-Neumann operator, 543, 546  
 Discontinuity arc, 533  
 Discontinuous system, 505  
     escaping vector field, 506  
     regularization, 507  
     sliding vector field, 506  
 Dissipation, McGehee coordinates, 382  
 Dissipative KAM theory, 32  
 Dissipative perturbations, 23, 25  
 Divisor of zeroes, 568  
 Double pendulum model of a planar biped, 528  
 Double well model  
     splitting, 399  
     symmetric double well, 402  
 Duffing oscillator, 22, 23  
 Dyachenko-Zakharov conjecture, 558–561  
 Dynamical Born–Oppenheimer approximation, 412  
 Dynamical systems, 185  
     linearizable dynamical system, 198  
 Dynamics, 218

## E

Eccentricity, 344  
 Effective Hamiltonian, 415  
 Egorov theorem, 404–406  
 Ehresmann fibration theorem, 307, 312  
 Eigenvalue, 105, 213  
 Eigenvectors, 51  
 Eight-shaped orbit,  $n$ -body problem, 366  
 Eight shaped three-body solution,  $n$ -body problem, 362  
 Einstein, Albert, 2  
 Elastodynamics, 245  
 Electrical networks, complex, 203  
 Electrodynamics, 1  
 Electromagnetic theory, 245  
 Electronic structure problem, 411  
 Elementary theory, 171–173  
 Eliasson, L.H., 6  
 Elliptic, 21  
 Elliptic restricted three-body problem (RTBP), 158  
 Energy band, 411  
 Energy equipartition, 300  
 Energy-momentum map, 149, 150  
 Energy shell, 36  
 Energy surface, 411  
 Epicycles, 1  
 Equation of motion, 49  
 Equinoxes, 347–349  
 Equivariant bifurcation theory, 187

- Equivariant orbits, 361
  - Ergodic hypothesis, 32
  - Euler–Hill symmetry, 373
  - Euler top, 321–322
  - Evolution, 490
  - Evolutionarily stable strategy, 495
  - Evolutionary game theory, 490
    - equilibria in, 494–495
    - perturbations of equilibria, 495–496
  - Exact energy, 54
  - Excitation spectrum, 478, 479
  - Exponentially small, 97
    - non-adiabatic transitions, 425–426
  - Exponential stability, 296
  - Extended phase space, 187
  - External parameters, 89
  - External symmetry, 199
- F**
- Fast convergent method, 248
  - Fast system, 507
  - Fermi, E., 2
  - Fermi ball, 471, 477, 482
  - Fermi liquid theory, 484
  - Fermion(s), 458–460, 469–472, 474, 477, 480–484
  - Fermionic Bogoliubov map, 473
  - Fermi–Pasta–Ulam–Tsingou (FPUT) recurrence, 575
  - Fermi systems, 9
  - Feynman diagrams, 461
  - Feynman graph, 3, 9, 461
  - Feynman–Hellmann theorem, 55–56
  - Fiber bundle, 203
  - Fibration, 307, 311
  - Finite-gap method, 568
  - Finite-gap potentials, 568
  - First-order approximation, 51
  - First-order corrections, 57
  - First-order singular partial differential equations, 218–219
  - Fisher’s theorem, 491
  - Fitness landscape, 490
    - evolution, 490
    - perturbation of equilibria, 492–493
    - stability of equilibria, 491–492
  - Floquet exponent, 36
  - Floquet form, 16
  - Floquet multiplier, 24
  - Floquet theory, 172, 175, 273
  - Flowbox theorem, 81
  - Fluctuations, 492
  - Fock space, 471–474, 482
  - Foliation, 308
  - Formal complete integrability, 113
  - Formal power series, 80, 121
  - Formal series, 2, 12
  - Fourier coefficients, 255, 342
  - Fourier integral operator, 403
  - Fourier multiplier, 550
  - Fourier representations, 462
  - Fourier series, 341, 342
  - Freak waves, 566
  - Frequency deviation, 142
  - Frequency-halving bifurcation, 37
  - Frequency spectrum, 7
  - Frequency vector, 122, 341, 346
    - Bryuno vector, 128
    - Diophantine vector, 128
  - Frobenius integrable, 308
  - Functional derivative, 231
  - Further normalization, 185
- G**
- Galactic dynamics, 203
  - Game theory, 493–494
  - Gap condition, 420–423, 425, 427, 428
  - Gaussian integrals, 457–458
  - Generalized creation and annihilation operators, 473
  - Generalized hip–Hops,  $n$ -body problem, 366
  - Generalized orbits, 377
  - Generalized solution,  $n$ -body, 377
  - Generalized Sundman–Sperling estimates,  $n$ -body problem, 381
  - General Lie algebra, 196
  - Generating functions, 250
  - Generic bifurcation, 510–512
  - Genericity, 96
  - Genetic locus, 492
  - Geometric singular perturbation (GSP), 507
  - Geometric singular perturbation theory (GSPT), 504, 512, 513, 515
  - Geometry of group action, 200
  - G-Equivariant minimizers,  $n$ -body problem, 384
  - Gevrey, 96–98
    - index, 219
    - regularity, 219
    - spaces, 211
  - Gevrey-2, 294
  - Global KAM theory, 34
  - Global Nekhoroshev theory, 301
  - Global perturbation theory, 80
  - Golden number, 175
  - Gradient, 231
  - Gram–Hadamard bound, 461
  - Gram–Hadamard inequality, 462
  - Grassmann integrals, 458–459
  - Grazing bifurcation, 521, 533
  - Grazing-sliding bifurcation, 533
  - Great multiplier, 294
  - Green function, 69, 70
  - Green’s theorem, 49
  - Gröbner basis, 100
  - Gross–Pitaevskii approximation, 476–480
  - Gross–Pitaevskii equation, 245
  - Gross–Pitaevskii scaling limit, 470, 471
  - Ground state, 470, 471, 476, 481, 482, 484
    - energy, 470, 476–480, 483, 484
  - Group theory, 64
  - Gyroscopic systems, 178, 180

**H**

Hadamard, 3  
 Hamiltonian, 90, 96–98, 177, 420–429, 438, 468, 470, 471, 475, 476, 478, 479, 481–483  
   chaos, 178  
   dynamical systems, 186  
   dynamics, 247  
   effective, 435–440  
   flow, 249–251  
   function, 174, 343, 347  
   Hopf bifurcation, 25  
   perturbations, 22  
   perturbation theory, 293, 294, 300, 302, 542  
   vector field, 186, 234, 309  
 Hamiltonian system, 15, 116, 177, 182, 203, 309  
   singular, 359  
 Hamilton–Jacobi equation, 262  
 Hamilton’s equations, 249  
 Hartree–Fock approximation, 476, 477, 480, 482  
 Hepp, K., 6  
 Hepp’s theorem, 3  
 Hessian matrix, 400  
 Heteroclinic orbit, 23  
 High-density scaling limit, 471–472  
 Higher-dimensional perturbation analysis, 534  
 Higher-order corrections, 54  
 Hilbert–Schmidt operator, 474  
 Hilbert space, 51, 52, 56, 62, 71, 252, 275, 420, 421, 426, 429, 461  
 Hill’s equation, 167, 168, 172, 174, 175  
 Hip–Hops,  $n$ -body problem, 366  
 Holomorphic function, 396  
 Holomorphic linearization, 97  
 Homoclinic orbit, 24  
 Homological equation, 81, 85, 141, 235, 294  
 Homological operator, 189  
 Hopf bifurcation, 15, 23, 34, 37, 39–41  
 Hopf fibration, 321  
 Hopf–Neimark–Sacker bifurcation, 42  
 Hubbard model, 461, 462  
 Hybrid zero dynamics approach, 526–527  
 Hyperbolic equilibrium, 16  
 Hyperbolic systems, 225  
 Hysteresis, 180

**I**

60-Icosahedral periodic minimizers, 370  
 Imperfect bifurcations, 441–443  
 Inductive hypothesis, 281  
 Infinite-dimensional torus, 237  
 Inhomogeneous equation, 69  
 Instability pockets, 167  
 Instability tongues, 167, 173  
 Integrability, 116  
 Integrable affine structure, 31  
 Integrable system, 16  
   integrable, 186  
 Integral (of motion), 15

Interaction number, 142  
 Intermediate normalization, 51, 54, 62  
 Interval arithmetic, 284  
 Intra-band dynamics, 414  
 Invariance relation, 201  
   elementary, 201  
 Invariant curves of planar diffeomorphisms, 25  
 Invariant KAM surfaces, 349–350  
 Invariant manifold, 114, 159–160, 202  
 Invariant tori  
   invariant KAM surfaces, 349–350  
   librational tori, for spin-orbit problem, 351  
   planetary problem, 352–353  
   rotational tori, for restricted three-body problem, 351–352  
   rotational tori, for spin-orbit problem, 350–351  
 Invariant torus, 122, 308  
 Inverse function theorem, 262, 279  
 Involution, 170  
 Iso-energetic KAM tori, 272  
 Isolated energy band, 414  
 Isosceles symmetry, 373  
 Isotropic oscillator, 147  
 Italian symmetry, 370

**J**

Jahn–Teller effect, 410  
 Jet transport, 160  
 Joint normal form, 194  
 Jordan blocks, 213, 216, 217, 219, 220, 222–226  
 Jordan canonical form (JCF), 213, 214

**K**

KAM stability, 299  
 KAM theory, 16, 158–159, 231, 328–329  
   Arnold’s scheme, 260–266  
   classical implicit function theorem, 278–279  
   description, 248  
   finite dimensional context, 249–252  
   infinite dimensional, 252–253, 274–277  
   Kolmogorov (*see* Kolmogorov theorem)  
   Kolmogorov set and Whitney smoothness, 271  
   lower dimensional tori, 267  
   Moser’s theorem, 266–267  
   motivation of, 248  
   non-degeneracy assumption, 272  
   PDEs, 248  
   perturbation of classical integrable systems, 272–273  
   physical stability estimates and break-down thresholds, 273–274  
   power series expansions, 271–272  
   spectral theory of Schrödinger operators, 273  
   topological trapping in low dimensions, 273  
 KAM torus, 349  
 Kato–Rellich theorem, 66  
 Kato’s quantum adiabatic theorem, 421–423  
 KdV equation, 277  
 Kepler’s equation, 13, 123

- Kepler's laws, 339
- Klein-Gordon equation, 275
- Kolmogorov, A.N., 3
- Kolmogorov-Arnold-Moser (KAM) theory, 18–20, 30, 120, 340
  - for gravity capillary water waves, 549–552
  - for pure gravity water waves in finite depth, 552–555
  - and quasi-periodic solutions, 541
- Kolmogorov nondegeneracy condition, 308
- Kolmogorov set, 260
- Kolmogorov's method, 5
- Kolmogorov's type condition, 318
- Kolmogorov theorem
  - estimates, 256–258
  - Hamiltonian functions, 253
  - iteration and convergence, 258–260
  - Kolmogorov transformation, 253–256
  - torus, 253
- Korteweg-de Vries (KdV) equation, 229, 233, 568
- L**
- Labeled tree, 124
- Lagrange symmetry, 374
- Lagrangian equilibrium, 300
- Lagrangian function, 232
- Lagrangian manifold, 280
- Lagrangian torus, 20
- Lamb shift, 2
- Landau, 4
- Landau-Zener Hamiltonian, 426
- Laplace, P.S., 2
- Laplace's limit, 13
- Lebesgue measure, 250, 260, 267
- Lee-Huang-Yang formula, 483
- Legendre polynomials, 344
- Level crossing, 65–66
- Levinson's theorem, 532
- Librational tori, for spin-orbit problem, 351
- Lie, 83
  - derivative, 83
  - exponentiation formula, 270
  - transforms, 177, 190
- Lie algebra, 88, 105, 220
  - arguments, 111–114
  - Nilpotent Lie algebra, 196
  - proof, 88
  - semisimple Lie algebra, 196
  - subalgebra, 88
- Lie groups, 187
  - derivative, 83, 89
  - group, 88
  - subgroup, 88
- Lie-Poincaré transformation, 191
- Limit set, 171
- Lindstedt
  - algorithm, 2
  - series, 6, 8, 11, 120
- Linear algebra, 193, 213
- Linear autonomous systems, 214
- Linear damping, 179
- Linear flow, 308
- Linearization, 81, 221
  - of complex maps, 29
  - continuous, 83
  - of dynamical system, 198
  - formal, 81
  - holomorphic, 97
  - smooth, 83
- Linear symmetries, 195
- Line symmetry, 372
- 3-Link planar bipedal robot, 526
- Liouville-Arnold's complete integrability, 312
- Liouville-Arnold theorem, 307
- Liouville integrable system, 35
- Liouville number, 217
- Lippmann-Schwinger equation, 70
- Lipschitz constant, 281
- Lipschitz functions, 268–270
- Local bifurcation, 25
- Localized coefficients, 236, 239
- Locally minimizing solution,  $n$ -body, 377
- Local trivialization, 307
- Local well posedness, 545
- Logarithmic type potentials,  $n$ -body problem, 383
- Low-density scaling limit, 471
- Lower dimensional tori, 33, 267–271
- Lower-order corrections, 52
- M**
- Macroscopic time scale, 411
- MACSYMA, 177
- Magnetohydrodynamics (MHD), 245
- Majorizing series, 109
- Many-body quantum mechanics, 469–470
- Marchal's P12-symmetry, 368
- Mathieu equation, 167, 168, 170, 172, 173, 176, 177
  - with viscous damping, 173
- Maximal KAM tori, 251
- Maximally superintegrable systems, 310, 319
- McGeer's model, 520
- McGehee coordinates, dissipation, 382
- Mean-field/semiclassical scaling limit, 477
- Melnikov conditions, 267
- Melnikov theory, 181
- Michel, Louis, 200
- Minimizer
  - $n$ -body problem, 366
  - planar equivariant, 368
  - three-dimensional equivariant, 368
- Modulation instability, 565, 566
- Molecular dynamics (MD), 409
- Moment map, 200
- Momentum mapping, 35
- Monodromy, 16, 433
  - classical, 446
  - fractional, 446
  - Hamiltonian, 438
  - quantum, 445



Moreau sweeping process approach, 534  
 Morrey, 3  
 Morse theory, 436  
 Moser's theorem, 266, 273  
 Moser's twist map theorem, 25  
 Mountain pass solution, 3-body problem, 376, 378  
 Moyal product, 404  
 Multiband adiabatic theory, 417  
 Multiple resonances  
   and quantum state density, 451–452  
 Multiplicative factor, 53  
 Multiplicity, algebraic, 213  
 Multiscale analysis, 1, 5–7, 12, 128  
 Multiscale problems, 6

## N

Nash equilibrium, 494–496, 498  
 Nash-Moser quadratic algorithm, 270  
 Nash-Moser scheme, 547, 553  
 $n$ -Body problem, 357  
   absence of collision, 383  
   basic definitions, 360  
   conservation laws, 381  
    $D_{4q} \times C_2$  group, 365  
    $D_6 \times C_3$  group, 367  
    $D_q \times C_2$  group, 366  
   eight shaped three-body solution, 362  
   generalized solution, 377  
   generalized Sundman–Sperling estimates, 381  
   G-equivariant minimizers, 384  
   locally minimizing solution, 377  
   Neumann boundary conditions, 384  
   rotating circle property, 363  
   Standard variation, 385  
 Nearly-integrable Hamiltonian, 16  
 Near resonances, 141  
 Neimark–Sacker  
   bifurcation, 40, 171  
 Nekhoroshev  
   estimates, 100  
   integrable degenerate systems, 312  
   stability, 299  
   theorem, 235, 324–326  
   theory, 20, 153, 296–300  
 Neumann boundary conditions, 233  
    $n$ -body problem, 384  
 Neumann series, 281  
 Newton method, 110  
 Newton's law, 2  
 Newton's tangent scheme, 279  
 Newton tangent method, 547  
 Nilpotent matrix, 216  
 Nilpotent perturbation, 212, 215–218, 222, 226  
 Noncommutative integrability, 310, 312  
 Non-degeneracy condition, 133  
 Nondegenerate case, 52  
   energy and eigenvectors, corrections to, 52–54  
   Feynman–Hellmann theorem, 55–56  
   Wigner's theorem, 54–55

Non Hamiltonian PDEs, 242–243  
 Nonlinear dynamics, 429  
 Nonlinear normal modes, 145  
 Nonlinear parametric excitation, 176–177  
 Nonlinear Schrödinger (NLS) equation, 229, 233, 237, 565  
 Nonlinear symmetries, 195  
 Non-planar choreography, 369  
 Non-planar minimizers, 376, 378  
 Non-planar symmetric orbit, 369  
 Non-resonant motion, 7  
 Nonresonant Nekhoroshev normal form, 327  
 Normal form, 79, 105, 141, 143, 170, 175, 177, 185, 193,  
   194, 212, 230, 231, 438, 541  
   Birkhoff, 92, 93, 97  
   for finite dimensional Hamiltonian systems, 234–236  
   for Hamiltonian PDEs, 236  
   for nonresonant Hamiltonian PDEs, 239–242  
   procedure, 81  
   for resonant Hamiltonian PDEs and consequences,  
     236–239  
   semi-local, 92  
   Takens, 92  
   theory, 84, 85, 88, 90, 93, 97, 99, 100, 215  
   truncation, 16  
   symmetry, 96  
 Normal form on an invariant manifold (NFIM), 114, 115  
 Normal frequency, 33  
 Normalization, 80, 85, 87, 143–144, 169, 171  
   quasi-resonant normalisation, 144  
 Normalized exact wave function, 54  
 Normalizing transformation, 93, 96, 97  
 Normal modes, 42, 49, 145  
 Number operator, 471, 472

## O

24-Octahedral periodic minimizers, 370  
 One degree of freedom, 22  
 Open quantum systems, 429  
 Operator-valued distribution, 473  
 Optimal exponent, 298  
 Orbit  
   generalized, 377  
   in generic position, 145  
   non-planar symmetric, 369  
 Orbital, 469  
 Orthogonal complement, 56  
 Orthogonality theorems, 63  
 Oscillator, harmonic, 435  
 Outstanding problems, 13

## P

Padé-approximants, 178  
 Painlevé's theorem, 379  
 Pairing density, 474  
 Parabolic equilibrium, 22  
 Paradiifferential calculus, 545  
 Parallel transport, 421, 422, 424

- Parametrically excited pendulum, 178
  - Parametrically forced oscillators, in sum resonance, 175–176
  - Parametric excitation, 168, 171, 176–177
    - of linear systems, 171–176
  - Parametric resonance, 167
  - Parametrization method, 160
  - Partial differential equations (PDEs), 218
    - finite dimensional Hamiltonian systems, normal form for, 234–236
      - Hamiltonian formalism for, 231–234
      - Hamiltonian PDEs, normal form for, 236
      - non Hamiltonian PDEs, 242–243
      - nonresonant Hamiltonian PDEs, normal form for, 239–242
      - resonant Hamiltonian PDEs, normal form for, 236–239
  - Particle, collisions in the  $n$ -body problem, 359
  - Particle–hole pair, 480, 481
  - Particle–hole transformation, 475, 481
  - Patches, 482, 484
  - Peregrine soliton, 567
  - Period-doubling bifurcation, 24
  - Periodic boundary conditions, 458
  - Periodic center-saddle bifurcation, 24
  - Periodic like trajectories, 230
  - Periodic orbits, 36
    - construction of, 353–354
    - libration in longitude of the Moon, 354
  - Periodic solution, 93
    - subharmonic, 94
  - Persistence, 80
  - Persistent, 80
    - property, 19
  - Perturbation, 79, 80, 211
    - averaging, 169
    - methods, 532
    - normalization of time-dependent vectorfields, 170–171
    - of periodic orbits, 24
    - Poincaré–Lindstedt series, 168–169
    - problem, 16, 80
    - resonance, 169
    - series, 66–68
  - Perturbation of continuous spectrum, 69
    - Born series and convergence, 71–73
    - scattering solutions and scattering amplitude, 69–71
  - Perturbation of point spectra
    - degenerate case, 56–61
    - nondegenerate case, 52–56
  - Perturbations of superintegrable systems, 323
    - applications, 329–332
    - KAM theory, 328–329
    - motions along symplectic leaves, 326–328
    - Nekhoroshev theorem, 324–326
    - semiglobal approach, 323–324
  - Perturbation theory (PT), 1, 7, 17–19, 22, 25, 30, 31, 49, 50, 57, 59, 61, 62, 74, 76, 185, 340, 528–531, 541, 576–581
    - classical perturbation theory, 341–345
    - Hamiltonian, 542
    - Lindstedt series, 7–8
    - multiscale analysis, 5–7
    - need of convergence proofs, 5
    - non convergent cases, 10–12
    - Poincaré’s theorem, 2–3
    - purpose of, 544
    - renormalization, 3–5
    - resonant perturbation theory, 345–349
    - and symmetry, 63–65
  - Perturbative determination of symmetries, 191
  - Perturbative expansions, 459–460
  - Perturbing function, 344
  - Phase locking, 180
  - Phase portrait, 435
  - Phase space, 16, 187
  - Pinched torus, 35
  - Pitchfork, 177
  - Planar 3-body problem, symmetry groups, 372
  - Planar Delaunay action-angle variables, 342
  - Planar equivariant minimizer, 368
  - Planar Hamiltonian systems, 22
  - Planar Kepler system, 320
  - Planar rimless wheel, 524–526
  - Planar symmetry groups, 372
    - 3-body problem, 372
  - Planetary problem, 352–353
  - Planetary system, 323
  - Plane wave, 471, 477, 480
  - Plasmon, 484
  - Poincaré, H., 2
  - Poincaré domain, 141, 216, 218
  - Poincaré–Dulac normal form (PDNF), 105–108, 189, 190
  - Poincaré–Dulac theorem, 224, 243
  - Poincaré–Lindstedt series, 168–169
  - Poincaré–Lyapounov–Nekhoroshev theorem, 198
  - Poincaré map, 19, 25, 520
    - 3D rimless wheel, 525
    - fixed points of perturbed, 527–528
    - perturbation theory, 528–531
    - for systems with impacts, 522–524
  - Poincaré morphism, 221, 222
  - Poincaré renormalized form, 199
  - Poincaré’s theorem, 2, 81, 83, 84
  - Poincaré transformation, 189
  - Poisson brackets, 234
  - Poisson manifold, 308
  - Polynomial Hamiltonian, 269
  - Polynomial map, 224
  - Power series, 105
    - expansions, 271
  - Primary resonance, 177
  - Prisoner’s dilemma, 497
  - $h$ -Pseudodifferential calculus, 403–404
    - elliptic symbol, 404
- Q**
- Quadratic Hamiltonian, 470, 475–476, 478, 479
  - Qualitative theory, 229
  - Quantities of interest, 470

- Quantum adiabatic theorem
  - Berry's connection and parallel transport, 423–424
  - exponentially small non-adiabatic transitions, 425–426
  - Kato's quantum adiabatic theorem, 421–423
  - space-adiabatic theorems, 426–428
  - super-adiabatic expansions, 424–425
- Quantum bifurcations (QB), 434, 435
  - imperfect bifurcations, 441–443
  - organization of bifurcations, 442–444
  - physical applications and generalizations, 452
  - simplest effective Hamiltonians, 435–440
  - symmetry, 440–442
  - two degree-of-freedom integrable systems, bifurcation diagrams for, 444–448
- Quantum correlations, 468, 477, 480, 481
- Quantum electrodynamics, 3
- Quantum field theory, 457, 459, 462
- Quantum mechanics, 49, 50, 245, 468–469
- Quantum particles, 458
- Quantum phase transition, 434
- Quasi-convexity, 307
- Quasi-degenerate case, 60–61
- Quasi-free approximation, 476–478
- Quasi-free state, 473–477, 483
- Quasi-homogeneous potentials,  $n$ -body problem, 383
- Quasi-linear system, 244
- Quasi-momentum, 568
- Quasi-periodic, 16, 167, 168, 175, 182
  - Bifurcations, 34, 38
  - Hamiltonian Hopf bifurcation, 98
  - Hopf bifurcation, 40–42
  - motion, 1, 2, 122, 158–159, 248
  - torus, 94
  - traveling water waves, 555
- Quasi-resonant normalization, 144
- R**
- Random phase approximation, 480, 481, 483, 484
- Rayleigh-Schrödinger (RS) perturbation theory, 50, 60, 61, 425
- Recursive solution, 192
- Redistribution of energy levels, 449
- Reduced dynamics, 520
- Reduced phase space, 149
- Reduced system, 80
- Reducibility scheme, 555
- Reduction, 80, 87, 92, 106, 434
- Reference frame, 343
- Reflection symmetry, 146, 177
- Regularization process, 507
- Regular perturbations, 66, 67
- Relative equilibria, 40
  - $n$ -body problem, 359
- Relativistic Thirring model, 581
- Relay systems, 504
- Remainder, 224
- Renormalization, 6
  - theory, 3, 12
- Renormalization group (RG), 1, 9, 12, 120
- Replicator equation, 492, 494
- Residual degeneracy, 57
- Resonance, 16, 81, 82, 87, 94, 142, 169, 218
  - adiabatic theorems for, 428–429
  - 2:1 resonance, 149
  - 2:1 resonance, 151
  - $k:l$  resonance, 145–146
  - relation, 190, 201
  - sporadic resonance, 201
  - symmetric 1:1 resonance, 146
  - symmetric 1:1 resonance, 149
  - tongues, 173
- Resonant lattice, 308
- Resonant monomial, 105, 108
- Resonant motion, 10
- Resonant normal form, 237
- Resonant perturbation theory
  - degenerate perturbation theory, 347
  - precession of the equinoxes, 347–349
  - resonant theory, 345–346
  - three-body resonance, 346–347
- Resonant points, 570
- Resonant scalar monomial, 201
- Resonant zone, 293
- Restricted-three body model, 273
- Restricted three-body problem (RTBP), 157, 351–352
- Resummation, 3, 9, 12, 131
  - resummed series, 12, 131
- Reversibility, water waves system, 549, 554
- Reversible, 89, 98, 170, 171
  - perturbations, 23
- Riemannian manifolds, 67
- Rimless wheel model, 520, 525
- Rogue waves, 565, 566
  - Cauchy problem for, 569
  - definition, 565
  - existence, 566
  - Fermi-Pasta-Ulam-Tsingou recurrence of, 575–576
  - finite-gap approximation, 569–572
  - future research, 581–582
  - leading order finite-gap solution, 571–572
  - periodic problem, NLS equation, 567–569
  - perturbation theory, 576–581
  - perturbed spectral data, 571
  - solution of Cauchy problem, 572–576
  - statistical aspects, 566
  - unperturbed spectral data, 570
- Rotating circle property (RCP), 363
- Rotational symmetry, 86
- Rotational tori
  - for restricted three-body problem, 351–352
  - for spin-orbit problem, 350–351
- Rotation number, 123
  - best approximant, 128
- Rotor, 178, 180
  - dynamics, 178–179
- Running couplings, 3, 12
- Rüssmann estimate, 257
- Rüssmann non-degeneracy condition, 272

## S

- Saddle, 24
- Saddle-node, 177
  - limit point bifurcation, 41
- Scale, 10
  - index, 12
- Scaling factor, 10
- Scaling laws, 3
- Scaling limit, 470–472, 477, 483, 484
- Scattering length, 476, 478, 479, 483
- Schrödinger equation, 56, 69, 70, 73, 409, 421, 470
- Second Melnikov conditions, 135
- Second-order equation, 62
- Second quantization, 468, 470, 472–473, 475
- Self-energy clusters, 10
- Self-energy resummations, 12
- Semiclassical approximation, 400
- Semiclassical limit, 391
- Semiclassical methods, 392
- Semiglobal approach, 323–324
- Semi-quantum limit, and reorganization of quantum bands, 449–451
- Semisimple Joint Normal Form (SJNF), 220
- Separatrices, 16
- Sets of analyticity, 114
- Ship dynamics, 203
- Siegel, K., 3
- Siegel–Bryuno lemma, 129
- Siegel domain, 216, 217, 222
- Siegel’s method, 6
- Šilnikov orbit, 181
- Simple, 80, 84
- Simple choreographies, minimizing properties, 375
- Simplified, 85
- Simplify, 80, 84
- Sin-Fourier series, 275
- Single frequency system, 294
- Single well potential, 398
- Singular hamiltonian systems, 359
- Singularity
  - focus-focus, 438
  - theory, 38
- Singular perturbation problem, 506–507
  - in 2D, 512–513
- Singular perturbations, 534
- Skew product system, 135
- Slater determinant, 470, 471, 474, 475, 477, 480–482
- Slow dynamics, 514
- Slowly varying attractors (SVAs), 580
- Slow system, 507
- Small denominator, 6, 8, 108, 141, 190, 230, 235, 294
- Small divisor, 95, 143
  - problem, 127, 340, 342
- Small-world networks, 497
- Smooth linearization, 83
- Snowdrift game, 499
- Sobolev space, 544, 550
- Solenoid, 36
- Soliton equations, 565
- Sophus Lie, 187
- Space three-body problem, 374
- Spatial Kepler system, 320
- Spatial perturbations, 496–498
- Spectral curve, 568
- Spectral projection, 420, 421, 424, 425, 428
- Spectrum, 215
- Spin-orbit problem
  - librational tori for, 351
  - rotational tori for, 350–351
- Splitting of separatrices, 35
- Spontaneous symmetry breaking, 200, 434
- Stability of planetary systems, 331
- Stable manifold, 16
- Stag hunt, 497
- Standard map, 123
- Standard variation,  $n$ -body problem, 385
- Steep function, 296
- Sternberg’s theorem, 83
- Stokes line, 392, 396
- Stokes rule, 397
- Stokes waves, 566
- Strange attractor, 171
- Stratification, 15
- Strong interactions, 1
- Strongly non resonant constant frequency system, 295
- Structural stability, 16, 80, 505
- Structure, 80, 88
  - preservation, 88, 94
- Summation at the smallest term, 295
- Summation rule, 2
- Sum resonance, 167, 171, 175, 179
- Sundman–Sperling estimates, generalized, 381
- Super-adiabatic evolution, 425
- Super-adiabatic expansions, 424–425
- Super-adiabatic Hamiltonian, 425
- Super-adiabatic subspace, 414
- Superexponential stability, 300
- Superintegrability, 318–319
- Superintegrable Hamiltonian system
  - definition, 309
  - dynamics, 317–318
  - Euler top, 321–322
  - geometric structure of integrable toric fibrations, 314–316
  - Hamiltonian case, 312–313
  - integrability, 311–312
  - maximally superintegrable systems, 319
  - noncommutative integrability, 312
  - perturbations, 323–332
  - planar Kepler system, 320
  - point in central force field, 321
  - spatial Kepler system, 320
  - symmetry and superintegrability, 322
  - symplectic nature, 313–314
- Superintegrable systems, 308
- Switching manifold, 521, 523
- Symbolic computation, 53
- Symmetric wave function, 469

Symmetry, 79, 106, 168, 185, 188  
 approximate symmetry, 193  
 axial, 90  
 axially, 86  
 and bifurcations, 440–442  
 breaking, 434  
 breaking, spontaneous, 434  
 and degeneracy, 62–66  
 generalized symmetry, 203  
 rotational, 86, 91  
 toroidal, 80, 86, 92, 95  
 torus, 85

Symmetry for systems in normal form, 197

Symmetry groups, 176, 361

core, 369

planar 3-body problem, 372

Symmetry of dynamical systems, 187

Symmetry reduction, 185, 200

Symplectic, 88, 292

Symplectically complete foliations, 319

Symplectic foliation, 308

Symplectic form, 32

Symplectic matrix, 250

## T

Takens–Bogdanov bifurcation, 23

Taylor expansion, 50

Taylor series, 341, 459

Taylor's formula, 256

12-Tetrahedral periodic minimizers, 370

Theorem

Flowbox, 81

Hartman–Grobman, 83

Noether, 80

Poincaré, 81, 83

Sternberg's, 83

Theory

Cantorized singularity, 98

KAM, 96, 99

normal form, 99

parametrized KAM, 99

perturbation, 79

singularity, 80, 94, 96

Three-body problem, 358

Three-body solution, eight shaped, 362

Three-dimensional equivariant

minimizer, 368

Time dependent perturbations, 73–76, 158

Time-dependent vectorfields, 170

Time scales, 498–499

Töplitz–Lipschitz hamiltonians, 276

Toroidal symmetry, 80, 86

Toroidal symmetry with small divisors, 95

Torus, 120

lower-dimensional, 133

maximal, 133

resonant, 133

symmetry, 85

Transcritical bifurcation, 150

Transformation, 86, 216

canonical, 91

equivariant, 89

normalizing, 80

s-preserving, 89

Transition probabilities, 54

Transversal, 81

Transversal linearization, 527

Tree, 7, 120, 124

cluster, 128

number, 8

renormalized, 132

self-energy cluster, 128

value, 8

Triviality conjecture, 4

Trivial monodromy, 317

Trivial symmetry, 372

Truncated expectations, 460–461

Truncation, 79, 86, 218

Tunneling between wells, 402–403

Turning points, 391

Twist condition, 28

Twist map, 28

Two-body problem, 358

## U

Ultimatum game, 501

Unlabeled tree, 124

Unperturbed system, 79

Unstable manifold, 16

## V

Vacuum, 472, 474, 481

Van-der-Waals equation, 2

Vector field, 17, 507–510

$C^0$  equivalent, 504

germ-equivalent, 505

Velocity potential, 542

Versal unfolding, 176

Virial series, 3

Volume preserving, 88

vector fields, 113

## W

Water waves

Dyachenko–Zakharov conjecture, 558–561

equation, 541

future developments, 561

long time existence for periodic, 556–561

nonlinear nature, 545

problem, 245, 542

quasi-periodic traveling, 555

Wave equation, 229

Wave function, 54

corrections, 53

Weierstrass, 3

Weyl operator, 474, 478

Weyl quantization, 404

Whitney smoothness, 98

Whitney umbrella, 176  
Wick rule, 458  
Wick theorem, 474–477  
Wigner crystal, 484  
Wigner's theorem, 54–55  
Wilson, K. G., 4  
WKB approximation, 393  
    connection formula, 395–398  
    semiclassical solutions in the classically allowed  
        region, 393–395  
    semiclassical solutions in the classically forbidden  
        region, 395

**Y**  
Yang–Mills theories, 5

**Z**  
Zeipel's theorem  
     $n$ -body, 379  
    structure in the collision set, 379  
Zeroth-order wave function, 57  
Zoll manifolds, 278